AD_____

Award Number: W81XWH-06-1-0611

TITLE: Identification of Novel Tumor Suppressor Genes in Human Breast Cancer Using Nonsense-Mediated mRNA Decay Inhibition (NMDI)-Microarray Analysis

PRINCIPAL INVESTIGATOR: Cameron Neil Johnstone, Ph.D.

CONTRACTING ORGANIZATION: University of Pennsylvania
Philadelphia, PA 19104

REPORT DATE: August 2007

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
01-08-2007	Final	15 Jul 2006– 14 Jul 2007

4. TITLE AND SUBTITLE

Identification of Novel Tumor Suppressor Genes in Human Breast Cancer Using Nonsense-Mediated mRNA Decay Inhibition (NMDI)-Microarray Analysis

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-06-1-0611

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Cameron Neil Johnstone, Ph.D.

Email: cameronj@mail.med.upenn.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University of Pennsylvania
Philadelphia, PA 19104

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This project sought to identify genes that harbor nonsense mutations in breast cancer cell lines that are commonly used as in vitro models in the study of breast cancer biology, with the ultimate aim of identifying novel tumor suppressor genes for sporadic breast cancer. We focused our efforts on the long arm chromosome 22 which is known to undergo LOH in primary breast tumors. Before the NMD-microarray strategy could be undertaken, we very thoroughly characterized chromosome 22q copy number and allelic imbalance in several breast cancer cell lines by integrating publicly available genetic data with empirical data derived from 317K single nucleotide polymorphism (SNP) arrays. MCF-7, T-47D, and MDA-MB-231 breast cancer-derived cell lines were selected for NMD-microarray analysis. Two different regimens for inhibiting NMD were then directly compared for their ability to specifically inhibit NMD while leaving wild-type transcripts unaffected. The improved second-generation NMD protocol was performed and we eagerly await the results of the Affymetrix GeneChip expression arrays so that interrogation of the expression data may begin.

15. SUBJECT TERMS
nonsense-mediated mRNA decay, microarray, quantitative RT-PCR, single nucleotide polymorphism array, deletion, allelic imbalance, p53, peroxisome proliferator activated receptor gamma, caffeine, Actinomycin D, emetine.

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
U	U	U	UU	35	19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

Introduction:

Eukaryotes have evolved sophisticated mechanisms to prevent the biosynthesis of mutant proteins that may produce potential deleterious effects on the overall fitness of the organism. One such mechanism is nonsense-mediated mRNA decay (NMD), a pathway conserved from yeasts to humans. Tumor suppressor genes (as well as genes mutated randomly) often harbor nonsense mutations in tumors, and the vast majority of their mutant transcripts are degraded by the NMD pathway to prevent the production/accumulation of truncated proteins. Recently, inhibition of the NMD pathway in cancer cells has been used to identify transcripts bearing nonsense mutations in cancer cells (1-5). The aim of this project was to: **(1)** identify breast cancer cell lines which harbor genetic alterations of chromosome 22q that are indicative of the presence of a potential tumor suppressor gene on this chromosome; and **(2)** to apply the NMD-microarray analysis strategy to identify candidate tumor suppressor genes on chromosome 22q bearing nonsense mutations in the aforementioned breast cancer cell lines.

Body:

(1) To identify the most appropriate breast cancer cell lines to use for this study, genetic alterations were assessed using complimentary approaches. Data from several different sources were integrated to arrive at a short list of breast cancer cell lines likely to harbor copy number loss, and/or allelic imbalance on the long arm of chromosome 22q, which could subsequently be interrogated in a more detailed analysis. Information was obtained from the following publically-available sources.

1) Microsatellite-based **allelic imbalance** data (7 loci): Catalogue of Somatic Mutations in Cancer (COSMIC). http://www.sanger.ac.uk/cgi-bin/genetics/CGP/genotyping/lohmap

2) Single nucleotide polymorphism (SNP) array-based **allelic imbalance** and **DNA copy number** data: Catalogue of Somatic Mutations in Cancer (COSMIC). http://www.sanger.ac.uk/cgi-bin/genetics/CGP/CGHviewer.cgi

3) Spectral karyotyping and **comparative genomic hybridization** (CGH) data: NCBI SKY/M-FISH and CGH Database. http://www.ncbi.nlm.nih.gov/projects/sky/ (6)

4) Published articles: N.C.B.I. PubMed. http://www.ncbi.nlm.nih.gov/sites/entrez?db=PubMed (7)

Analysis and integration of the aforementioned genetic and genomic data revealed four breast cancer cell lines with abnormal chromosome 22q structure. **Figure 1** displays the array-based comparative genomic hybridization (CGH) profiles for chromosome 22q in MCF-7, T-47D, and BT-474 cells (7). Statistically significant DNA copy loss was observed for MCF-7 and T-47D cell lines, but not for BT-474 cells. Additional information was obtained from the COSMIC SNP arrays (8, 9). First-generation Affymetrix 10K SNP arrays were used to generate the information contained in the COSMIC database. Despite representing a major improvement over CGH arrays, the average spacing between SNPs in the 10K array was likely to still be too small to detect changes involving only a small interval. Therefore, to obtain a more detailed profile, genomic DNA was prepared from these three breast cancer cell lines and also MDA-MB-231 and Hs578T breast cancer cells (data not shown) and applied to Illumina's 317K HumanHap300-Duo Genotyping BeadChip (http://www.illumina.com/pages.ilmn?ID=153). This BeadChip contains over 318,000 tag SNPs derived from Phase I HapMap data resulting in a median inter-SNP spacing of approximately 5 kb. Importantly, the BeadChip contains a higher density of tag SNPs in regions of the genome that are within 10 kb of a gene or in evolutionarily conserved regions. These data confirmed that none of the four cell lines contained normal diploid chromosome 22 (see **Figure 2A-2E**).

(2) Nonsense-mediated mRNA decay is initiated when the translational machinery encounters a premature stop codon (introduced by point mutation or frameshift mutation) in an exon preceding the terminal exon. The ribosome senses the inappropriate location by interaction with components of the spliceosome that remained associated with the

downstream exon-exon junction following splicing of the intervening intron. This then triggers a kinase signaling cascade, which culminates in the recruitment of endonucleases and exonucleases that degrade the offending transcript. Therefore, transcripts containing truncating mutations can be stabilized by treating cancer cells with inhibitors of translation such as emetine or puromycin (1). A strategy for identification of potential tumor suppressors was developed involving two separate gene profiling comparisons. In the first comparison, cells are treated with emetine alone (100 μg/ml) or vehicle. In the second comparison, cells are treated with emetine (100 μg/ml) plus Actinomycin D (2 μg/ml) or Actinomycin D alone (1). As inhibition of translation produces a non-specific stress response involving *de novo* transcription, gene transcription was also blocked in the second comparison using Actinomycin D. Therefore, transcripts that truly bear truncating mutations should be upregulated in both comparisons after gene expression profiling of the drug-treated cells. Indeed, this strategy (termed **g**ene **i**dentification by **N**MD **i**nhibition; GINI) was used to identify *EPHB2* as a novel tumor suppressor gene in prostate cancer (3).

However, Ionov and colleagues correctly pointed out that the second microarray comparison is technically challenging due to the low overall mRNA levels resulting from a four-hour treatment with Actinomycin D (5). To overcome this, a modified strategy was developed termed GINI-2 [see attached publication; (5)]. In this strategy, the first comparison is very similar to that in GINI-1, whereby cells treated with an inhibitor of NMD are compared to vehicle alone-treated cells. However, the second comparison entails treatment with NMD inhibitor plus Actinomycin D or Actinomycin D alone **after** a pre-treatment of the cells with the NMD blocker. Pre-treatment of cells with the NMD inhibitor allows the accumulation of both nonsense-mutated and stress response genes. Following this, transcripts containing a true nonsense mutation(s) should be degraded rapidly when Actinomycin D is applied in the absence of ongoing NMD blockade (5). This should occur less so when transcription is inhibited in conjunction with continuing blockade of NMD (**Figure 3**). The other major difference in GINI-2 is that the NMD pathway is blocked by inhibition of a protein kinase whose activity is required for the initiation of NMD (10). Specifically, activity of the phosphatidylinositol 3' kinase-like

enzyme hSMG-1 is inhibited using caffeine (10 mM), rather than employing a global inhibition of translation with emetine (11).

As a result of these protocol modifications, GINI-1 and GINI-2 were compared (**Figure 4**) for their ability to detect a truncating *TP53* mutation in PC-3 prostate cancer cells (mutant) and MCF-7 breast cancer cells (wild-type). As expected, treatment of cells with emetine alone for 8h or caffeine alone for 4h resulted in an increase of p53 mRNA levels. In GINI-1, treatment with Actinomycin D and emetine (4h) resulted in a 7.8-fold increase in expression compared with emetine alone (4h). Similarly, following a 4h pre-treatment with caffeine in GINI-2, subsequent treatment with Actinomycin D and caffeine (4h) lead to a 6.2-fold increase in p53 mRNA levels over Actinomycin D alone (4h). Surprisingly, there was little difference in p53 mRNA abundance after incubation with Actinomycin D regardless of whether pre-incubation with caffeine occurred (**Figure 4**). In MCF-7 cells, both sets of drug treatments in both GINI-1 and GINI-2 resulted in very modest increases in p53 gene expression (**Figure 5**).

In addition to the trials using the *TP53* gene, potential off-target effects of both treatment regimens was also assessed by measuring responses of the *PPARG* gene (wild-type in both PC3 and MCF-7 cell lines). Treatment with emetine alone or caffeine alone yielded modest increases in total PPARG mRNA levels in PC3 cells (**Figure 6**). However, GINI-2 outperformed GINI-1 in the second set of drug treatments. The combination of Actinomycin D and caffeine produced a 2.1-fold increase in PPARG mRNA versus control, whereas Actinomycin D plus emetine resulted in a 4.3-fold increase over control, therefore suggesting that GINI-2 may have fewer off-target effects than GINI-1. In MCF-7 cells, the GINI-2 regimen also perturbed PPARG mRNA levels less than did GINI-1 (**Figure 7**).

Primarily due to an apparent milder effect on wild-type mRNA levels, GINI-2 was selected for use in the NMD-microarray analysis in the current study. Hence, three breast cancer cell lines (MCF-7, T-47D, and MDA-MB-231) with chromosome 22q alterations (see above) were subjected to GINI-2 in duplicate followed by RNA isolation using Trizol reagent, thus resulting in 8 independent RNA samples for each cell line. The 24 RNA samples were submitted for gene expression analysis at the Penn Microarray

Core Facility (http://www.med.upenn.edu/microarr/) on Affymetrix U133_Plus_2 gene chips and the results of this analysis are eagerly anticipated.

Key Research Accomplishments:

- Genome-wide assessment of DNA copy number and allelic imbalance in MCF-7, T-47D, BT474, MDA-MB-231, and Hs578T breast cancer cell lines using Illumina 317K HumanHap300-Duo Genotyping BeadChip.
- Comparison of GINI-1 versus GINI-2 protocol for NMD-microarray study with selection of GINI-2.

Reportable Outcomes:

- Genome-wide DNA copy number and allelic imbalance data in MCF-7, T-47D, BT474, MDA-MB-231, and Hs578T breast cancer cell lines (and cultured normal human skin fibroblasts) using Illumina 317K HumanHap300-Duo Genotyping BeadChip.
- Data from the 24 samples subjected to NMD-microarray analysis will be submitted to the Gene Expression Omnibus (GEO) repository for microarray data (http://www.ncbi.nlm.nih.gov/geo/).

Conclusion:

This project sought to identify genes that harbor nonsense mutations in breast cancer cell lines that are commonly used as in vitro models in the study of breast cancer biology, with the ultimate aim of identifying novel tumor suppressor genes for sporadic breast cancer. We focused our efforts on the long arm chromosome 22 which is known to undergo LOH in primary breast tumors. Before the NMD-microarray strategy could be undertaken, we very thoroughly characterized chromosome 22q copy number and allelic imbalance in several breast cancer cell lines by integrating publicly available genetic data with empirical data derived from 317K SNP arrays. MCF-7, T-47D, and MDA-MB-231 cell lines were subjected to the GINI-2 NMD protocol and we await the results of the Affymetrix expression arrays so that interrogation of the expression data may begin.

References:

1) Noensie E.N., and Dietz H.C. A strategy for disease gene identification through nonsense-mediated mRNA decay inhibition. Nat. Biotechnol. (2001) **19:**434-439.

2) Wolf M., Edgren H., Muggerud A., Kilpinen S., Huusko P., Sørlie T., Mousses S., and Kallioniemi O. NMD microarray analysis for rapid genome-wide screen of mutated genes in cancer. Cell Oncol. (2005) **27:**169-173 (Review)

3) Huusko P., Ponciano-Jackson D., Wolf M., Kiefer J.A, Azorsa D.O., Tuzmen S., Weaver D., Robbins C., Moses T., Allinen M., Hautaniemi S., Chen Y., Elkahloun A., Basik M., Bova G.S., Bubendorf L., Lugli A., Sauter G., Schleutker J., Ozcelik H., Elowe S., Pawson T., Trent J.M., Carpten J.D., Kallioniemi O.P., and Mousses S. Nonsense-mediated decay microarray analysis identifies mutations of EPHB2 in human prostate cancer. Nat. Genet. (2004) **36:**979-983

4) Ionov Y., Nowak N., Perucho M., Markowitz S., and Cowell J.K. Manipulation of nonsense mediated decay identifies gene mutations in colon cancer cells with microsatellite instability. Oncogene (2004) **23:**639-645

5) Ivanov I., Lo K.C., Hawthorn L., Cowell J.K., and Ionov Y. Identifying candidate colon cancer tumor suppressor genes using inhibition of nonsense-mediated mRNA decay in colon cancer cells. Oncogene (2007) **26:**2873-2884

6) Roschke A.V., Tonon G., Gehlhaus K.S., McTyre N., Bussey K.J., Lababidi S., Scudiero D.A., Weinstein J.N., and Kirsch I.R. Karyotypic complexity of the NCI-60 drug-screening panel. Cancer Res. (2003) **63:**8634-8647

7) Pollack J.R., Sørlie T., Perou C.M., Rees C.A., Jeffrey S.S., Lonning P.E., Tibshirani R., Botstein D., Børresen-Dale A.L., and Brown P.O. Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors. Proc. Natl. Acad. Sci. USA. (2002) **99:**12963-12968

8) Forbes S., Clements J., Dawson E., Bamford S., Webb T., Dogan A., Flanagan A., Teague J., Wooster R., Futreal P.A., and Stratton M.R. COSMIC 2005. Br. J. Cancer (2006) **94:**318-322

9) Bamford S., Dawson E., Forbes S., Clements J., Pettett R., Dogan A., Flanagan A., Teague J., Futreal P.A., Stratton M.R., and Wooster R. The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. Br. J. Cancer (2004) **91:**355-358

10) Brumbaugh K.M., Otterness D.M., Geisen C., Oliveira V., Brognard J., Li X., Lejeune F., Tibbetts R.S., Maquat L.E., and Abraham R.T. The mRNA surveillance protein hSMG-1 functions in genotoxic stress response pathways in mammalian cells. Mol. Cell (2004) **14:**585-598.

11) Usuki F., Yamashita A., Higuchi I., Ohnishi T., Shiraishi T., Osame M., and Ohno S. Inhibition of nonsense-mediated mRNA decay rescues the phenotype in Ullrich's Disease. Ann. Neurol. (2004) **55:**740-744

Appendices:

Ivanov I., Lo K.C., Hawthorn L., Cowell J.K., and Ionov Y. Identifying candidate colon cancer tumor suppressor genes using inhibition of nonsense-mediated mRNA decay in colon cancer cells. Oncogene (2007) **26:**2873-2884

Oncogene (2007) 26, 2873–2884

www.nature.com/onc

ORIGINAL ARTICLE

Identifying candidate colon cancer tumor suppressor genes using inhibition of nonsense-mediated mRNA decay in colon cancer cells

I Ivanov, KC Lo, L Hawthorn, JK Cowell and Y Ionov

Department of Cancer Genetics, Roswell Park Cancer Institute, Buffalo, NY, USA

Inhibition of the nonsense-mediated decay (NMD) mechanism in cells results in stabilization of transcripts carrying premature translation termination codons. A strategy referred to as gene identification by NMD inhibition (GINI) has been proposed to identify genes carrying nonsense mutations. Genes containing frameshift mutations in colon cancer cell line have been identified using a modified version of GINI. To increase the efficiency of identifying mutant genes using GINI, we have now further improved the strategy. In this approach, inhibition of NMD with emetine is complemented with inhibiting NMD by blocking the phosphorylation of the hUpf1 protein with caffeine. In addition, to enhance the GINI strategy, comparing mRNA level alterations produced by inhibiting transcription alone or inhibiting transcription together with NMD following caffeine pretreatment were used for the efficient identification of false positives produced as a result of stress response to NMD inhibition. To demonstrate the improved efficiency of this approach, we analysed colon cancer cell lines showing microsatellite instability. Bi-allelic inactivating mutations were found in the *FXR1*, *SEC31L1*, *NCOR1*, *BAT3*, *PHF14*, *ZNF294*, *C19ORF5* genes as well as genes coding for proteins with yet unknown functions.
Oncogene (2007) **26,** 2873–2884. doi:10.1038/sj.onc.1210098; published online 6 November 2006

Keywords: nonsense-mediated decay; mutations; caffeine; GINI

Introduction

It is widely accepted that the transformation of colon epithelial cells results from the activation of oncogenes as well as the inactivation of tumor suppressor genes controlling cell proliferation and survival (Vogelstein and Kinzler, 2004). In colon cancer cell lines with microsatellite instability (MSI) (Aaltonen *et al*., 1993; Ionov *et al*., 1993; Thibodeau *et al*., 1993), the high overall mutation frequency caused by the inactivation of DNA mismatch repair (MMR) genes (Fishel *et al*., 1993; Leach *et al*., 1993) results in the mutational inactivation of tumor suppressor genes. Consequently, these cell lines provide a valuable tool for identifying colon cancer related tumor suppressor genes. According to the two-hit hypothesis of tumorigenesis (Knudson, 1971), inactivating mutations affecting both alleles of a gene suggest a tumor suppressor function for the mutated gene.

Even though the genes mutated in colon tumors with MSI are often different from those mutated in MSI-negative tumors, the tumorigenic pathways affected by these mutations are often the same. For example, the absence of p53 mutations in colon cancers with MSI is often accompanied by mutations in the *BAX* (Rampino *et al*., 1997) or *P300* genes (Ionov *et al*., 2004a), both of which are involved in the p53 signaling pathway. Another example involves the transforming growth factor (TGF)β signaling pathway. In contrast to MSI-positive tumors, colon cancers without MSI rarely have inactivating mutations in the *TGFβ* receptor type II gene, although they frequently show mutations in *SMAD4*, a downstream effector of TGFβ signaling (Salovaara *et al*., 2002). This observation implies that identifying genes containing bi-allelic inactivating mutations in colon cancer cell lines with MSI will identify potential tumor suppressor genes or tumorigenic pathways related to malignant transformation. This observation is reinforced if the expression of these genes is also lost in primary tumor samples.

Gene inactivating events frequently result from nonsense or frameshift mutations, which create premature translation termination codons (PTC). Such mutations frequently elicit a rapid degradation of the mutant mRNA through the nonsense-mediated mRNA decay (NMD) pathway (Holbrook *et al*., 2004; Maquat, 2005; Weischenfeldt *et al*., 2005). NMD is an mRNA surveillance pathway, which prevents the production of truncated proteins that result mainly from the products of aberrant splicing. NMD is activated when multi-protein exon junction complexes (EJC) assemble on the sites of exon/exon junctions that are not removed from the spliced mRNA after the pioneer round of translation (Ishigaki *et al*., 2001). These EJCs are removed from the spliced mRNA during the movement of translated mRNA through the ribosome. When the translation

Correspondence: Dr Y Ionov, Department of Cancer Genetics, Roswell Park Cancer Institute, Elm and Carlton Streets, Buffalo, NY 14263, USA.
E-mail: Yurij.Ionov@RoswellPark.org
Received 15 June 2006; revised 21 August 2006; accepted 27 September 2006; published online 6 November 2006

machinery stalls on a PTC located more than ~50 nucleotides upstream of the last exon/exon junction, the unremoved EJCs downstream of the PTC recruit the Upf protein complex, which initiates the degradation of the PTC-containing mRNA transcripts (Holbrook *et al*., 2004).

Inhibition of translation has been shown to inhibit NMD and result in the accumulation of PTC-containing mRNA transcripts (Ishigaki *et al*., 2001). The increase in mRNA levels produced by inhibition of NMD can be detected using gene expression microarrays (Noensie and Dietz, 2001). The strategy of gene identification by NMD inhibition (GINI) provides an opportunity for the genome-wide identification of genes containing inactivating mutations (Noensie and Dietz, 2001). Using cDNA microarrays, Noensie and Dietz (2001) have demonstrated that blocking translation with emetine, for example, results in an elevated mRNA content for genes known to contain bi-allelic inactivating mutations. Although several genes mutated in colon and prostate cancer cell lines have already been identified (Huusko *et al*., 2004; Ionov *et al*., 2004b; Rossi *et al*., 2005), the efficiency of identifying mutations using the GINI strategy is still low, because too many genes that do not contain mutations also show mRNA increases following inhibition of NMD as a result of a stress response. This complication often makes it difficult to select candidate genes for sequencing analysis.

To improve the efficiency of identifying mutant genes using GINI, we evaluated different methods of NMD inhibition in this study. As a result, we have developed a modified version of GINI analysis (termed GINI2) that is more efficient in distinguishing those genes containing PTC from false positives that show mRNA increases due to a stress response to NMD inhibition. When we applied this GINI2 version to the analysis of colon cancer cell lines, a significantly increased number of genes containing bi-allelic inactivating mutations were identified. Some of these genes were also frequently mutated in primary colon tumors.

Results

Inhibition of NMD with caffeine can be used to detect NMD-mediated degradation of PTC-containing mRNA transcripts

We have shown previously (Ionov *et al*., 2004b) that increases of mRNA levels following inhibition of NMD can result from increased transcription of genes owing to a stress response to emetine. To overcome this problem, we developed a strategy to detect stabilization of mutant mRNA by combining the inhibition of NMD with the inhibition of transcription using actinomycin D. Although this approach has proved successful in identifying inactivating mutations in colorectal cancer cell lines (Ionov *et al*., 2004b), the analysis was compromised by the overall reduction in the levels of mRNA transcripts because of actinomycin D treatment. As a result, detectable hybridization signal for ~50% of the genes present on the microarray was lost. Thus, genes that might have been mutated, but which were expressed at only moderate to low levels, were excluded from detection.

To overcome this limitation, we have modified our method of simultaneous inhibition of NMD and transcription (Ionov *et al*., 2004b). An additional step has been introduced into our original protocol (Ionov *et al*., 2004b), which involves inhibiting NMD for several hours before blocking either transcription alone or transcription together with NMD. Inhibiting NMD before blocking both transcription and NMD should result in higher basal amounts of mutant mRNA, and consequently should result in stronger microarray hybridization signal for mutant genes. The predicted changes in mRNA levels for mutant and stress response genes following these different protocols of NMD inhibition are shown schematically in Figure 1.

This strategy was tested using the LS180 and RKO colon cancer cell lines, which we have previously shown to carry mutations in the *TGFBR2* and *P300* genes, respectively. These mRNAs are degraded by NMD (Ionov *et al*., 2004b) in these cell lines. Surprisingly, no significant difference in mRNA levels from the mutant genes were seen after incubation in the presence of actinomycin D, with or without emetine, following the initial pretreatment with emetine (data not shown). This was possibly due to the fact that emetine levels remain high within the cells, even after an extensive washing procedure. We, therefore, investigated using a different pharmacological inhibitor of NMD.

Sequential phosphorylation and dephosphorylation of the hUpf1 protein by the human suppressor with morphogenetic effect on genitalia-1 (hSMG-1) phosphatidylinositol 3-like kinase and protein phosphatase 2A (PP2A), respectively, is required for the initiation of NMD-induced mRNA degradation (Ohnishi *et al*., 2003; Brumbaugh *et al*., 2004). Moreover, a pharmacological block of NMD using caffeine or wortmannin, which inhibit hSMG-1 phosphorylation, has been reported previously (Usuki *et al*., 2004). We therefore analysed whether inhibition of either hSMG-1 kinase or PP2A phosphatase affects the stability of the mutant *P300* and *TGFBR2* genes in RKO and LS180 cells. Incubation with either caffeine (an inhibitor of hSMG-1 kinase), or okadaic acid or cantharidin (inhibitors of PP2A phosphatase) were shown using reverse transcription–polymerase chain reaction (RT–PCR) analysis (Figure 2a) to increase the amounts of the mutant *P300* and *TGFBR2* mRNAs in RKO and LS180 cells, respectively. Importantly, at the doses required to inhibit NMD, incubation with caffeine resulted in only minimal cell death, compared with okadaic acid and cantharidin. We therefore replaced emetine with caffeine in our GINI strategy. Northern blot analysis (Figure 2b) confirms that a significant increase of mutant *TGFBR2* mRNA was present following incubation of LS180 cells with caffeine for 4.5 h. Next, we tested whether pretreatment with caffeine would affect NMD activation following caffeine withdrawal. Northern blot analysis (Figure 3) demonstrates that pretreatment with caffeine, but not with emetine, results in significant

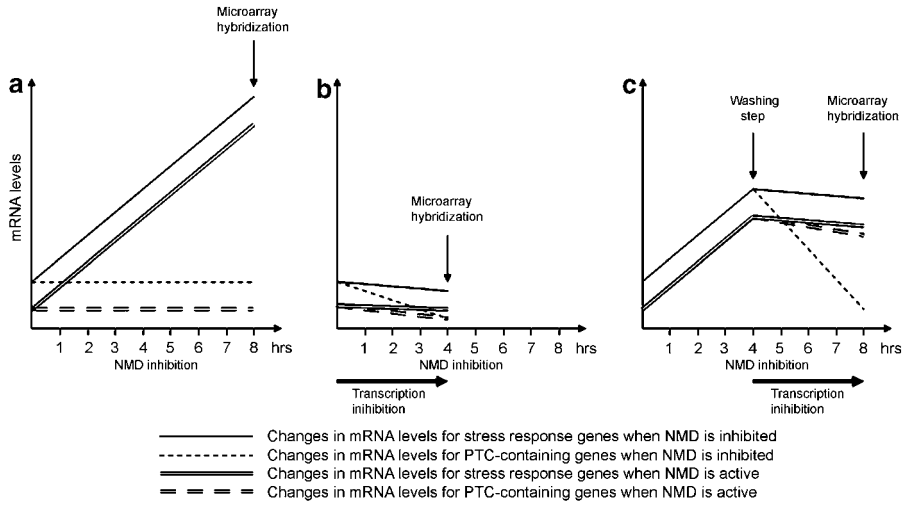


Figure 1 Theoretical depiction of changes in mRNA levels for stress response and PTC-containing genes following different protocols to inhibit NMD. (**a**) Continuous treatment of cells with an inhibitor of NMD results in increase of mRNA levels for PTC-containing genes, as well as for stress response genes. Microarray hybridization analysis cannot distinguish between the two causes of these increases: the enhanced mRNA synthesis due to stress response or the decreased mRNA degradation in case of PTC-containing genes. (**b**) When transcription is prevented using actinomycin D, even though NMD is blocked, the levels of PTC-containing transcripts and stress response genes are too low to be efficiently detected on microarrays. Changes in mRNA levels induced by NMD inhibition detected by hybridization analysis for many genes are within the 'hybridization noise', and are not reliable. (**c**) Blocking NMD for 4 h allows the accumulation of both PTC-containing transcripts and stress response gene transcripts to levels where they can be detected on the expression arrays. When transcription is blocked after this initial accumulation, the degradation of PTC-containing transcripts will occur quickly if the NMD blocking agent is removed, or will be delayed when NMD is sustained. The degradation of stress response transcripts after blocking transcription, on the other hand, does not depend on NMD. As a result, significant mRNA differences will be detected on the expression array for PTC-containing genes but not for the stress response genes.

differences in mutant *TGFBR2* mRNA levels in LS180 cells following either inhibition of transcription together with NMD, or inhibition of transcription only. This result illustrates that caffeine pretreatment can be used as an alternative to emetine to enhance the detectability of the microarray hybridization signals for mutant transcripts following inhibition of transcription with actinomycin D.

Identifying genes containing inactivating mutations in colon cancer cell lines using inhibition of NMD and Affymetrix genechip analysis

We used Affymetrix Genechip hybridization to measure the mRNA profile alterations produced by two alternative strategies of NMD inhibition, which are illustrated in Figure 1a and c. In the first method (Figure 1a), emetine was used to inhibit NMD as described in the original GINI protocol (Noensie and Dietz, 2001). The second method is a modification of our previously described (Ionov *et al*., 2004b) protocol for simultaneous inhibition of NMD and transcription, which is illustrated in Figure 1b. For convenience, we refer to the original emetine treatment as GINI1 and the modified caffeine protocol as GINI2. In the GINI2 protocol, caffeine was used to first inhibit NMD, thus allowing for an accumulation of mutant mRNAs in the cells (Figure 1c). The caffeine was then removed and the cells were treated further with either actinomycin D alone or actinomycin D and caffeine together. RNA from these experiments was then analysed using the U133Plus2.0 arrays.

It has been shown that the genes mutated in MSI-positive cells are generally different from those mutated in MSI-negative cells (Ionov *et al*., 1993; Konishi *et al*., 1996; Lengauer *et al*., 1998). Thus, to identify mutated genes in MSI (+) LS180 cells, we compared the GINI1 and GINI2 results from this cell line with those obtained using MSI (−) SW480 cells, assuming that different genes would be mutated in these two cell lines. Total RNA was isolated from both LS180 and SW480 cells treated according to either GINI1 or GINI2 protocols, as well as from untreated cells, and the mRNA levels were analysed using Affymetrix U133Plus2.0 GeneChip hybridization. The hybridization data were then compared between treated and untreated cells and the signal log ratios (SLRs), which equal Log_2 of fold changes of hybridization signal intensities produced by drug treatments, were calculated using the Affymetrix Microarray Suite version 5.0. The row data represented as .CEL files for GINI1 and GINI2 analyses for SW480 and LS180 cells have been submitted to GEO databases (accession number GSE5486).

Candidate genes for subsequent sequence analysis were selected according to three arbitrarily chosen parameters: (1) show a $SLR = Log_2$ (fold change) $\geqslant 1.7$ in LS180 cells and an $SLR \leqslant 1$ in SW480 following emetine treatment, (2) show $SLR \geqslant 1$ in LS180 cells following the GINI2 protocol and (3) show the intensity of hybridization signal for untreated LS180 cells lower

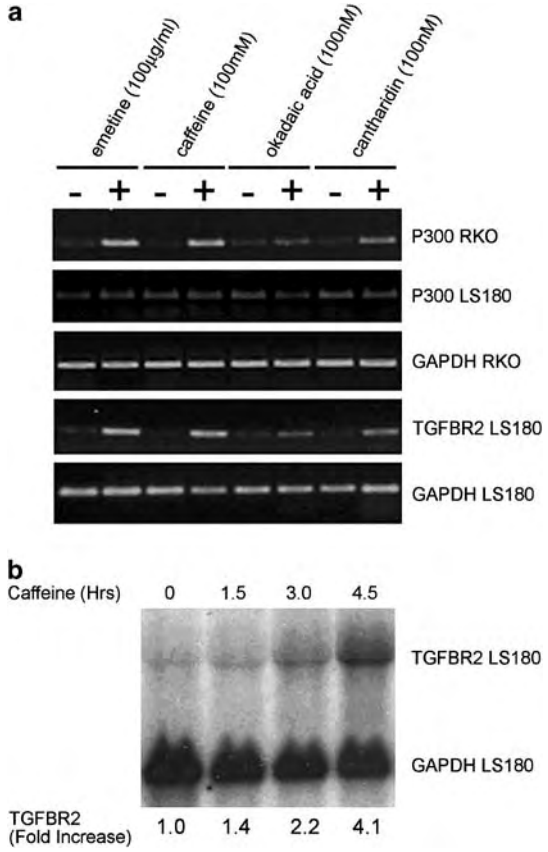


Figure 2 Accumulation of mutant *P300* and *TGFBR2* mRNAs in RKO and LS180 cells, respectively, following different approaches of NMD inhibition. (**a**) RT–PCR analysis demonstrates that, similar to incubation with emetine, incubation with 10 mM caffeine as well as with 100 nM okadaic acid or 100 nM cantharidin results in increased amounts of mutant (in RKO cells) but not wild-type (in LS180 cells) *P300* mRNA, as well as mutant *TGFBR2* mRNA (in LS180 cells). Total RNAs for the RT–PCR analysis were prepared following 3-h incubation with the indicated drugs. (**b**) Quantitative analysis of *TGFBR2* mRNA levels using Northern blotting. PhosphorImaging demonstrates significant accumulation of mutant *TGFBR2* mRNA in LS180 cells following 4.5-h incubation with 10 mM caffeine.

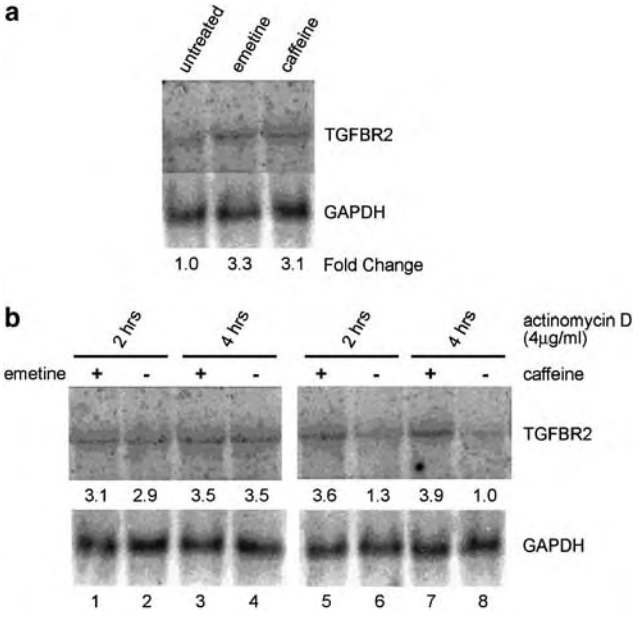


Figure 3 Blocking mRNA synthesis with actinomycin D, following inhibition of NMD with caffeine, but not with emetine enhances detection of NMD-dependent degradation of mutant *TGFBR2* mRNA in LS180 cells. (**a**) Northern blot analysis demonstrates that after 3-h incubation with either emetine (100 μg/ml) or caffeine (10 mM) increased amount of mutant *TGFBR2* mRNA in LS180 cells compared with untreated cells is observed. (**b**) (Left) Blocking transcription with actinomycin D following emetine pretreatment results in similar amounts of *TGFBR2* mRNA when NMD is either inhibited with emetine (lanes 2 and 4) or not (lanes 1 and 3). Blocking transcription with actinomycin D (Right) following caffeine treatment results in significantly different amounts of *TGFBR2* mRNA when NMD is either inhibited with caffeine (lanes 6 and 8) or not (lanes 5 and 7). Quantification (fold change) was performed using PhosphorImaging.

than that for the untreated SW480 cells. This follows as we expected that the levels of mutant transcripts in LS180 cells should be lower than corresponding wild-type levels in SW480 cells as a result of NMD degradation of mutant transcripts. This approach has proved effective in the past in our analysis of prostate cancer cell lines (Rossi *et al*., 2005). To further minimize the false discovery rate produced by hybridization 'noise', we excluded those genes which were scored as 'absent' by the Affymetrix analysis tool, either in the emetine treated LS180 cells or the untreated control SW480 cells. Of the approximately 55 000 probe sets representing approximately 20 000 unique genes on the Affymetrix U133Plus2 gene expression array, 15 candidate genes were identified (Table 1a) using these criteria. Sequencing the cDNA from these genes in LS180 cells has identified bi-allelic inactivating mutations in 7 genes (Table 1a).

To select candidates for sequencing we used crude differences in intensity to calculate the changes in transcript levels. As the number of false positives generated by this analysis was still relatively high (eight out of 15 candidates), we tested whether using additional statistical analysis could improve the efficiency of candidate selection. To do this, we used an adaptation of Robust Multi-Chip Average (RMA) (Irizarry *et al*., 2003), with specific correction for GC biases known as GC-RMA (Wu *et al*., 2004). When this statistical analysis was applied to the GINI1 and GINI2 row data, four genes from the original list were eliminated as well as one gene (CTNND1), which had been shown to contain bi-allelic mutations. This gene was located at the limit of the threshold cutoff, which supports the idea of reanalysing the data with more liberal thresholds. As a result of GC-RMA analysis, therefore, the efficiency of identifying mutant genes was greater than 50% (six mutant genes out of 10 candidates as seen in Table 1b). We also used GC-RMA analysis to test the efficiency of selecting candidate genes following GINI1 or GINI2 protocols separately. Neither GINI1 nor GINI2 alone achieve the same efficiency of analysis of combined GINI1 and GINI2 protocols. A possible explanation

Table 1a Candidate genes for sequencing in LS180 cells

Probe set ID	*Gene title*	*Gene symbol*	*LS180 emetine SLR (change)*	*LS180 C(A+C)/CA SLR (change)*	*LS180 untreated signal (detection)*	*SW480 untreated signal (detection)*	*SW480-emetine SLR (change)*
219121_s_at[a]	**RNA-binding motif protein 35A**	***RBM35A***	**3.6(I)**	**1.1(I)**	**143.8(A)**	**727.4(P)**	**0.8(I)**
225897_at	**Myristoylated alanine-rich protein kinase C substrate**	***MARCKS***	**3.5(I)**	**6.7(I)**	**182.4(A)**	**540(P)**	**0.6(NC)**
201669_s_at	**Myristoylated alanine-rich protein kinase C substrate**	***MARCKS***	**3.1(I)**	**5.2(I)**	**158.6(P)**	**1162.6(P)**	**0.7(I)**
222810_s_at	RAS protein activator like 2	*RASAL2*	3.1(I)	1.5(I)	33.4(A)	854.5(P)	0.5(NC)
208944_at	**Transforming growth factor, beta receptor II (70/80 kDa)**	***TGFBR2***	**2.5(I)**	**2.2(I)**	**1974.7(P)**	**2132.9(P)**	**−1.2(D)**
218522_s_at	**BPY2 interacting protein 1**	***BPY2IP1***	**2.2(I)**	**3.8(I)**	**339.6(P)**	**548(P)**	**−0.2(NC)**
201637_s_at	**Fragile X mental retardation, autosomal homolog 1**	***FXR1***	**2.1(I)**	**2.1(I)**	**368(P)**	**6638.8(P)**	**0.1(NC)**
206777_s_at	Crystallin, beta B2	*CRYBB2*	2(I)	1.3(I)	267.2(A)	1080(P)	0.2(NC)
211240_x_at	**Catenin (cadherin-associated protein), delta 1**	***CTNND1***	**1.9(I)**	**1(I)**	**832.4(P)**	**1730.1(P)**	**0.4(NC)**
222760_at	Zinc finger protein 703	*ZNF703*	1.9(I)	1.3(I)	1856.6(P)	2655.4(P)	0.4(NC)
224326_s_at	Polycomb group ring finger 6	*PCGF6*	1.9(I)	1.1(I)	302.8(P)	698.5(P)	0.1(NC)
225548_at	Shroom	*SHRM*	1.9(I)	1.5(I)	573.7(P)	1764(P)	0.8(I)
235521_at	Homeo box A3	*HOXA3*	1.8(I)	3(I)	270.6(P)	1480.2(P)	−0.3(D)
204391_x_at	Tripartite motif-containing 24	*TRIM24*	1.7(I)	1.8(I)	902.5(P)	1211.9(P)	0.9(I)
229231_at	Leucine rich repeat containing 37B	*LRRC37B*	1.7(I)	1.1(I)	735.1(P)	1030.9(P)	0.9(I)

LS180 as well as SW480 cells were treated with emetine and in addition LS180 cells underwent C(C+A)/CA treatment (See Methods). Changes in mRNA amounts produced in cells by treatments recorded as signal $\log_2$ ratio (SLR) as well as absolute hybridization signal intensities in untreated cells were analysed using the Affymetrix GeneChip Human Genome U133Plus2.0 array. In the table, the numbers for the SLRs are accompanied with comments in parenthesis indicating statistically significant increased (I), decreased (D) or not changed (NC). Absolute signal intensities numbers are accompanied in parenthesis with the comments indicating present (P) or absent (A). Selected genes had to satisfy the following parameters: (1) SLR $\geqslant$1.7 and SLR $\geqslant$1 in LS180 cells following emetine and C(C+A)/CA treatment, respectively, (2) SLR $<$1 in SW480 cells following emetine treatment, (3) absolute signal intensity corresponding to untreated LS180 cells less than that corresponding to untreated SW480 cells, (4) absolute signal intensity corresponding to SW480 cells indicated as present (P) should be $>$500. [a]**In bold** are the genes with identified biallelic inactivating mutations.

Table 1b Reduction of the candidate gene list from Table 1a after additional filtering using GCRMA normalized data

Probe set ID	*Gene title*	*Gene symbol*	*LS180 emetine SLR (change)*	*LS180 C(A+C)/CA SLR (change)*	*SW480-emetine SLR (change)*
219121_s_at*	**RNA-binding motif protein 35A**	***RBM35A***	**5.51**	**1.10**	**0.72**
225897_at	**Myristoylated alanine-rich protein kinase C substrate**	***MARCKS***	**4.68**	**3.99**	**−0.09**
201669_s_at	**Myristoylated alanine-rich protein kinase C substrate**	***MARCKS***	**3.93**	**5.52**	**0.75**
208944_at	**Transforming growth factor, beta receptor II (70/80 kDa)**	***TGFBR2***	**2.75**	**2.56**	**−1.50**
218522_s_at	**BPY2 interacting protein 1**	***BPY2IP1***	**2.99**	**2.92**	**−0.46**
201637_s_at	**Fragile X mental retardation, autosomal homolog 1**	***FXR1***	**2.60**	**1.36**	**0.05**
206777_s_at	Crystallin, beta B2	*CRYBB2*	3.89	1.31	−0.26
222760_at	Zinc finger protein 703	*ZNF703*	2.85	1.31	0.35
225548_at	Shroom	*SHRM*	1.77	1.42	0.54
235521_at	Homeo box A3	*HOXA3*	2.59	3.42	−0.62

The signal log ratios (SLRs) are generated from the pairwise subtraction between signal log values after GCRMA normalization using 'gcrma' library from Bioconductor (www.bioconductor.org).

why GINI2 alone is not as efficient as in the combination with GINI1is that, despite the caffeine pretreatment, the subsequent incubation with actinomycin D results in lower levels of cellular mRNA. As a result, diminished signal intensities of hybridization are seen, effectively decreasing the 'signal to noise ratio' and consequently identifying a larger number of false positives.

Our results demonstrate that analysing changes in mRNA levels produced using both GINI1 and GINI2 protocols treatment to inhibit NMD is a highly efficient way to detect mutant genes. By analysing changes in mRNA levels following the GINI2 protocol, it was possible to distinguish genes that were stabilized by NMD inhibition from false positives that show mRNA increase due to increased transcription induced by a stress response. Semiquantitative RT–PCR analysis (Figure 4) illustrates how false positives produced by stress response to emetine treatment in LS180 cells can be successfully identified using the GINI2 protocol.

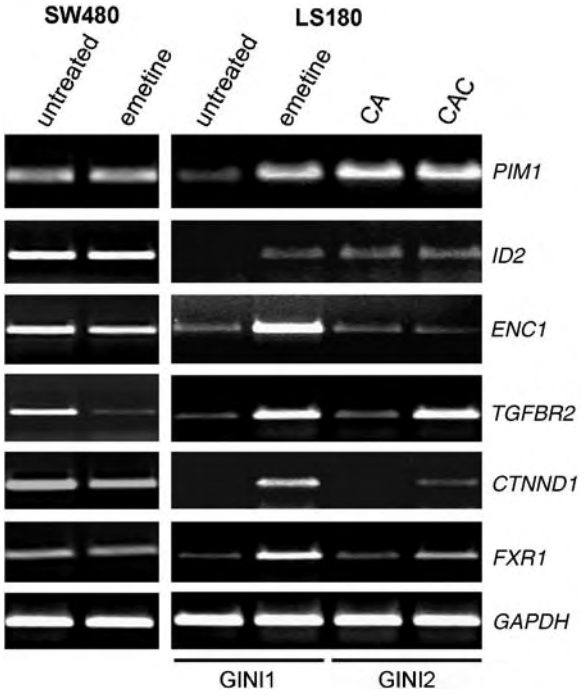


Figure 4 Selection of candidate genes for sequencing using analysis of changes in mRNA levels produced using both GINI1 and GINI2 protocols. These examples illustrate the strategy for prioritizing genes for sequencing following GINI protocols. RT–PCR analysis shows mRNA increases for six genes in LS180 but not in SW480 cells following emetine treatment (GINI1). As the level of mRNA for these genes is lower in LS180 than in SW480 cells, this suggests that basing only on GINI1 analysis all six genes would be selected as candidates for sequencing. GINI2 analysis, however, demonstrates the mRNA increases only for the *TGFBR2*, *CTNND1* and *FXR1* genes containing mutations but not for the *PIM1*, *ID2* and *ENC1* genes not harboring mutations in LS180 cells. CAC is the abbreviation for the pretreatment with caffeine followed by treatment with both actinomycin D and caffeine to block both transcription and NMD. CA is the abbreviation for the pretreatment with caffeine followed by treatment with actinomycin D alone.

RT–PCR analysis is shown for six different genes, none of which are mutated in the SW480 cell line. Emetine treatment in the GINI1 strategy in these cells does not result in an increased level of mRNA from these genes. In the LS180 cells, the mRNA levels for these genes in untreated cells is lower than in SW480 cells, which could be due either to normal variation between the cell lines or as a consequence of NMD. These genes, therefore, pass the first requirement in the prioritization. Next, the fact that emetine treatment of LS180 results in increased mRNA levels for these genes suggests this is due to either a stress response or lack of degradation by NMD, fulfilling the second requirement. Using the GINI2 strategy, no mRNA levels increases are seen for three of these genes (*PIM1*, *ID2*, *ENC1*), suggesting that they are stress response genes rather than carrying mutations. For the other three genes (*TGFBR2*, *CTNND1* and *FXR1*), increases in mRNA levels are seen after both GINI1 and GINI2 protocols, and sequencing analysis shows that they all carry mutations.

To identify mutated genes in the MSI-negative HT29 colon cancer cells, we compared mRNA profile changes following NMD inhibition in HT29 cells with that in LS180 cells. Using the same procedures and selection algorithm, we identified only four candidate genes in this comparison. This group included the *MADH4* gene, which has previously been reported to be homozygously mutated in HT29 (Woodford-Richens *et al*., 2001). No mutations were found in the remaining three candidates. When we lowered the stringency of the analytical filter, by using a lower numerical 'cutoff' for the GINI1 or GINI2 fold change thresholds, the number of candidates increased, but still no additional mutant genes were identified in HT29 cells. Furthermore, we could not identify PTC-generating mutations in MSI-negative SW480 colon cancer cells when LS180 or HT29 cells were used as controls.

The genes containing bi-allelic inactivating mutations identified so far in MSI-positive colon cancer cell lines using inhibition of NMD and Affymetrix genechips are described in Table 2. The prevailing mutation type in this study was the deletion or insertion of a single nucleotide in short-coding mononucleotide repeats. Such sequences represent mutational hotspots in cells with inactivated MMR (Markowitz *et al*., 1995; Rampino *et al*., 1997) and facilitated the analysis of these genes in MSI-positive primary colon tumors. To establish whether the genes identified as mutated in the colon cancer cell lines were also mutated in primary colon tumors showing MSI, the mononucleotide repeat-containing regions in these genes were sequenced in 36 tumors obtained from Oncomatrix Inc. (www.oncomatrix.com). Table 2 shows the mutation frequencies in primary colon tumors with MSI for the genes containing coding mononucleotide repeats longer than seven repetitive units which are mutated in cell lines. All analysed genes, except *SMAP-1*, were found to contain mutations in primary tumors with the frequencies largely proportional to the length of the coding microsatellites. However, some genes had higher mutation frequencies than other genes with the same length of the coding microsatellite. Moreover, the $(A)_8$ repeat of the *RBM35A* gene is mutated with higher (9/36, 25%) frequency than the $(A)_9$ repeat of the *RHAMM* gene, which is mutated in six of 36 (17%) tumors, suggesting the stronger selective pressure for the *RBM35A* gene inactivation in primary MSI-positive tumors. The absence of mutations in primary tumors for the *SMAP-1* gene, encoding for the stromal membrane-associated protein 1, suggests that the mutations in the HCT116, SW48 and RKO cells could occur during *in vitro* propagation.

Discussion

The original GINI strategy (GINI1) to identify genes containing nonsense mutations has been used successfully in proof-of-principle experiments to identify genes with known mutations in cell lines. Although it has also been used successfully to identify novel mutant genes in

Table 2 Genes mutated in colon cancer cell lines identified using GINI

Gene	*Gene symbol*	*Mutation*	*Cell line with gene mutated*	*Function*
RNA-binding motif protein 35A	*RBM35A*	DelA in $(A)_8$ nt 1528–35	LS180	RNA binding
Myristoylated alanine-rich protein kinase C substrate	*MARCKS*	DelA in $(A)_{11}$ nt 454–64	LS180, LoVo	Actin filament binding
BPY2 interacting protein 1	*BPY2IP1*	DelC in $(C)_6$ nt 588–93, Del AG in $(AG)_4$ nt 1447–54	LS180	Interacts with natural paclitaxel-like microtubule stabilizer
Fragile X mental retardation, autosomal homolog 1	*FXR1*	DelA in $(A)_8$ nt 364–71	Ls180	RNA binding
Catenin (cadherin-associated protein), delta 1	*CTNND1*	TGG→TGA nt 1431, DelTG in $(TG)_2$ nt 940–4	LS180	Cell-cell adhesion
Ring finger protein 43	*RNF43*	DelC in $(C)_6$ nt 343–48, DelG in $(G)_7$ nt 1969–75, DelTG in $(TG)_3$ nt 891–5	HCT116, SW48	Protein ubiquitination
SEC31-like 1 (Saccharomyces cerevisiae)	*SEC31L*	DelA in $(A)_9$ nt 1376–84	HCT116	ER to Golgi transport
Nuclear receptor co-repressor 1	*NCOR1*	DelC in $(C)_4$ nt 5494–7, InsA in $(A)_6$ nt 1568–73, DelTG in $(TG)_2$ nt 2463–6	RKO, HCT116	Regulation of transcription
mutS homolog 3 (Escherechia coli)	*MSH3*	DelA in $(A)_8$ nt 1141–8	RKO, HCT116	DNA mismatch repair
Hyaluronan-mediated motility receptor	*RHAMM*	DelA in $(A)_9$ nt 1990–8	RKO	Cell motility
PHD finger protein 14	*PHF14*	DelA in $(A)_7$ nt 530–6	HCT116	Regulation of transcription, DNA-dependent
Zinc finger protein 294	*ZNF294*	DelA in $(A)_{11}$ nt 1597–1607	Lovo, HCT116, RKO	Protein ubiquitination
Stromal membrane-associated protein 1	*SMAP-1*	Del A in $(A)_{10}$ nt 425–434	HCT116, SW48, RKO	Regulation of GTPase activity
HLA-B associated transcript 3	*BAT3*	DelC in $(C)_8$ nt 1180–7	LoVo	Implicated in the control of apoptosis and regulating heat shock protein
Beta-2-microglobulin	*B2M*	Del CT in $(CT)_4$ nt 37–44	LoVo	MHC class I receptor activity

Abbreviations: Del, deleted; ER, estrogen receptor; Ins, inserted; nt, nucleotides.

different systems, the problems associated with distinguishing stress response genes from mutated genes led to extensive unnecessary sequencing of genes that turned out not to carry mutations even despite custom filtering of the data. Even the use of actinomycin D to suppress *de novo* mRNA synthesis provided limited advantages, as the overall reduction in mRNA levels as a result of this treatment prevented detection of events related to moderately expressed genes. In this report, we describe a second-generation version of GINI analysis (GINI2), which uses caffeine as the inhibitor of NMD and incorporates a pretreatment step to boost overall mRNA levels. Using this strategy, in combination with the original GINI protocol, we have identified a series of genes, which are mutated in colon caner cell lines with MSI. In many cases, these genes have also been shown to be mutated in primary colon cancer tumors with MSI. Thus, the GINI2 approach has proved a significant development in the search for colon cancer-related genes. Using a modified GINI strategy, we have analysed five MSI-positive colon cancer cell lines and identified 13 novel genes that are mutated in colon cancers with MSI.

As arbitrarily chosen cutoff thresholds were used to select candidate genes for sequencing analysis, we have likely missed some genes harboring NMD-activating mutations in the cell lines analysed. For example, our series of mutant genes does not include *BAX*, *RIZ*, *PTEN*, *RAD50*, *CASPASE-5*, *IGFIIR*, *MSH6*, *DNA-PKCs* and many others that have been reported to be mutated at high frequencies in colorectal cancers with MSI. It is possible that many of these genes do not contain bi-allelic frameshift mutations in the cell lines analysed. The *BAX* gene, however, is known to harbor mutations in both alleles in LoVo and LS180 cells (Rampino *et al*., 1997), but has not been identified as the target for sequencing using our analysis. This fact is in agreement with recent publication describing differential NMD of mutated mRNAs in MMR-deficient colorectal cancers (El-Bchiri *et al*., 2005). The cause for the difference in the efficiency of NMD for the degradation of mutant mRNAs between individual genes is unknown. It is possible that the extent of the increase in mutant mRNA levels produced by NMD inhibition depends on the half-life of the corresponding wild-type mRNA transcripts. For example, if non-mutant mRNA for a hypothetical gene is degraded regardless of NMD immediately after the first round of translation, then the inhibition of NMD for the same gene harboring PTC will not result in accumulation of the mutant transcript. If this suggestion is true, then identifying mutant genes using GINI analysis for the transcripts with short half-lives will be less efficient than for the transcripts with longer half-lives. Loosening the stringency of the

analytical filter, by using a lower numerical 'cutoff' for the fold change thresholds in GINI1 or GINI2 will increase the number of candidates for sequence analysis, which might increase the number of mutant genes identified. This may, however, also result in a higher number of false positives, in turn reducing the efficiency of identifying mutant genes.

The GINI strategy can efficiently identify only those mutations, which affect both alleles of a gene and result in PTC inactivating gene function. Such mutations suggest possible tumor suppressor function of these genes. Identifying inactivating mutations in both alleles of the *CTNND1* gene in LS180 cells (Figure 5a) illustrates the efficiency of GINI for the identification of candidate

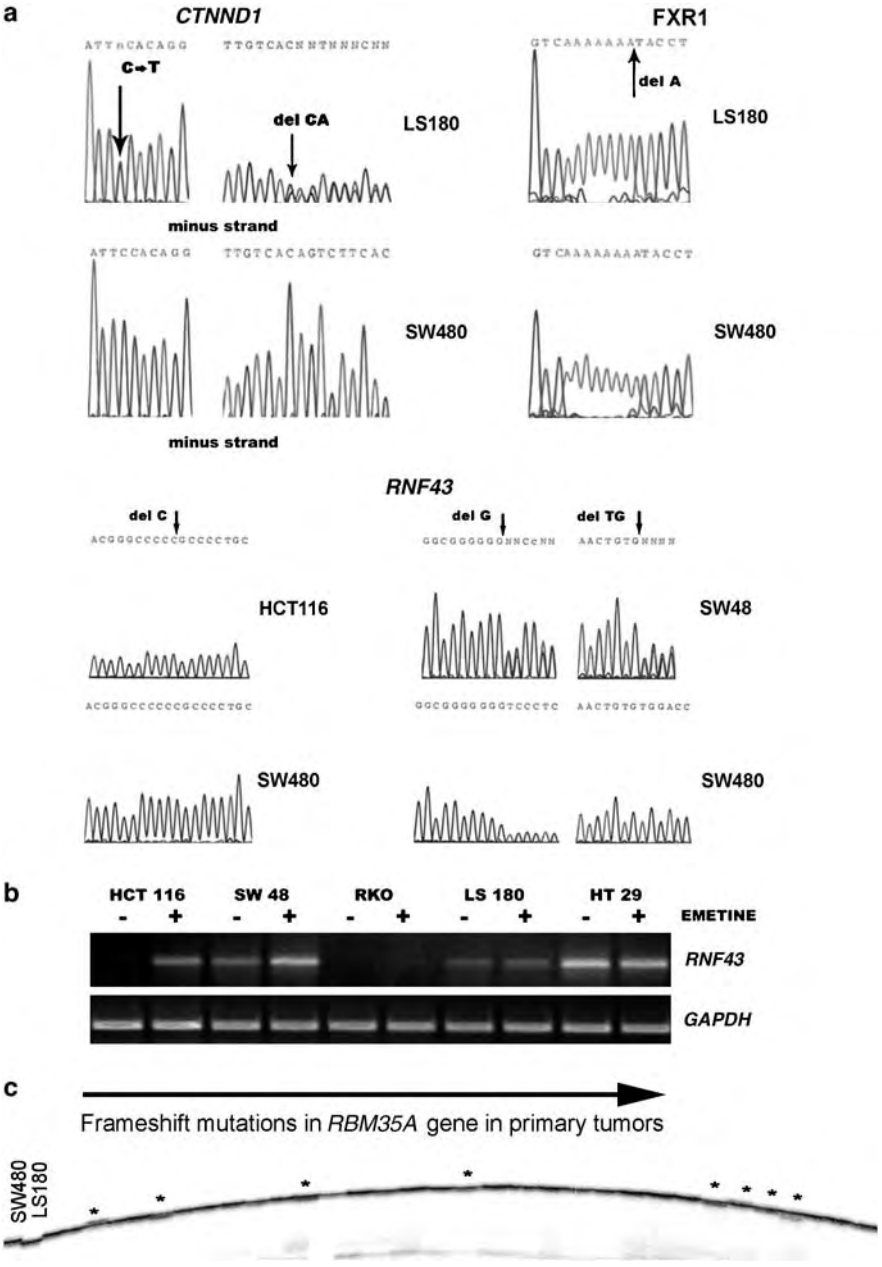


Figure 5 Mutation analysis of candidate genes selected using GINI. (**a**) Sequencing analysis identified two heterozygous mutations in LS180 cells in the *CTNND1* gene (one is a G to A substitution resulting in a TGA stop codon and the other is TG deletion in a $(TG)_2$ repeat (shown on the minus strand sequencing chromatogram). A homozygous deletion of one adenosine in *FXR1* gene is seen, and a homozygous deletion of one cytosine and two heterozygous mutation (a deletions of a G in a $(G)_7$ and TG in a $(TG)_3$ repeats) are seen in *RNF43* in HCT116 and SW48 cells, respectively. Wild-type sequences of the corresponding normal region from SW480 cells are shown as controls. (**b**) RT–PCR analysis of *RNF43* mRNA levels in both untreated and emetine treated colon cancer cells. Cell lines HCT116 and SW48, with identified inactivating mutations in the *RNF43* gene, show mRNA increases following inhibition of NMD with emetine, unlike LS180 and HT29, which do not have *RNF43* mutations. Cell line RKO does not express RNF43 mRNA. (**c**) Denaturing sequencing gel electrophoresis shows shifts in mobility for nine PCR-amplified fragments indicating frameshift mutations. Frameshift mutations in the *RBM35A* gene in colon tumors with MSI (indicated by the *) sign were detected by autography of PCR-amplified genomic DNA region containing the $(A)_8$ tract.

tumor suppressor genes. *CTNND1* has been also found to be mutated by others in SW48 MSI-positive colon cancer cells and the tumor suppressor function of this gene had been supported by functional studies (Ireton *et al*., 2002; Thoreson and Reynolds, 2002).

The majority (12/13) of the mutations in colon cancer cell lines identified using GINI analysis occurred in coding microsatellite sequences in the genes involved. On a background of inactivated MMR, it is possible that these mutations could occur as a consequence of the high overall instability of microsatellite DNA (bystander effect), rather than being tumor cell promoting events. Clearly, the only way to distinguish between these possibilities is to perform functional studies to determine whether they play a role in the development of the malignant phenotype. GINI analysis provides the candidate genes to perform these analyses. There have, however, been some estimates of the mutation frequencies in coding microsatellites of the genes contributing to tumor development (target genes) (Boland *et al*., 1998; Perucho, 1999a, b; Duval *et al*., 2001, 2002; Zhang *et al*., 2001; Suzuki *et al*., 2002; Woerner *et al*., 2003), which appears to depend largely on the length of the repeat (Kunkel *et al*., 1994; Chen *et al*., 1995). According to statistical analyses describing the relationship between the mutation frequency and the relevance to tumor development, the frequency of the mutations in coding microsatellites in bystander genes (not related to tumorigenesis) are no higher than the frequency of mutations in non-coding microsatellites of the same length and type (Duval *et al*., 2002; Woerner *et al*., 2003). The frequency of mutations seen in many non-coding repeats that are less than nine mononucleotides, has been shown to be lower than 6% (Suzuki *et al*., 2002) and so the mutations seen in the *RBM35A* gene, for example, are less likely to be due to a bystander effect, as this gene was mutated in 25% of primary tumors with MSI (Table 2 and Figure 5).

To our knowledge, there are no studies describing the frequency of mutations at non-coding microsatellites that are less than eight nucleotides long. During our screen for mutations using PCR products derived from genomic regions containing mononucleotide repeats in the regions flanking exons in over 100 sequencing analyses, we have never encountered mutations in these sequences. We suggest, therefore, that bi-allelic frameshift mutations occurring in the short-coding microsatellites of *BPY2IP1*, *NCOR1* and *RNF43* genes (Table 2) are unlikely to be due to background instability of coding microsatellite repeats. Moreover, each allele of these genes was inactivated by different heterozygous mutations, indicating the absence of a particularly unstable repetitive sequence in the coding DNA in these genes. Instead, these mutations may be considered to provide a selective advantage as a result of gene inactivation.

The protein encoded by *BPY2IP1*, also known as cell death inducer *C19ORF5*, is a hyperstabilized microtubule-specific binding protein, which causes mitochondrial aggregation and cell death (Liu *et al*., 2005c). Its involvement in tumor suppression is suggested by its interaction with natural paclitaxel-like microtubule stabilizer and candidate tumor suppressor *RASSF1A* (Liu *et al*., 2002, 2005a–c), which is frequently silenced by promoter hypermethylation in colon tumors with MSI (Oliveira *et al*., 2005).

Although the role of the *NCOR1* gene in suppression of colon tumorigenesis has not been reported, there is evidence for its involvement in the development of breast cancer. *NCOR1* associates with estrogen receptor-α (ER-α) preventing tamoxifen from stimulating proliferation in breast cancer cells. This effect is achieved through repression of a subset of target genes involved in ER-α function and cell proliferation (Keeton and Brown, 2005). It has also been shown that low NCOR1 expression in breast tumors was associated with a significantly shorter relapse-free survival (Girault *et al*., 2003) and that patients with high *NCOR1* expression levels have a better prognosis than those with low expression (Zhang *et al*., 2005). Mutational inactivation of this gene in colon cancer cell lines also suggests a role for *NCOR1* in colon cancer.

The RING finger protein 43 (*RNF43*), which shows homozygous deletion of one nucleotide in a $(C)_6$ repeat in HCT116 cells, and two heterozygous mutations in other parts of the gene in SW48 cells (Figure 5a), does not express mRNA in RKO cells as shown by RT–PCR analysis. MSI in RKO cells is known to be the consequence of *hMLH1* gene silencing due to promoter hypermethylation (Veigl *et al*., 1998). It is possible, therefore, that the loss of expression of the *RNF43* gene in RKO cells may also be due to an epigenetic mechanism. Loss of gene expression in one cell line, and the mutational inactivation in the others, argues in favor of the selective advantage provided by inactivation of the *RNF43* gene. The presence of a RING finger domain in the protein sequence can be associated with the ubiquitin-protein ligase activity of the gene product, suggesting a possible involvement in the mechanism of cell cycle progression as well as in the regulation of the expression of other genes by initiating the degradation of their protein products (Fang *et al*., 2003).

In addition to *RNF43* and *NCOR1*, several other genes that contain bi-allelic inactivating mutations in colon cancer cell lines may affect the gene expression profiles of cancer cells. The *RBM35A* gene, which contains an RNA-binding motif may be involved in mRNA splicing or degradation. *ZNF294*, which is mutated in all of the MSI-positive colon cancer cell lines analysed, also contains a RING finger motif which, like *RNF43*, may regulate the expression of other genes through its ubiquitin-protein ligase activity. *FXR1*, which is mutated in LS180 cells, has recently been shown to post-transcriptionally regulate the expression of genes containing AU-rich elements within the 3′-untranslated regions of their mRNA. This regulation modulates translational efficacy and mRNA stability (Garnon *et al*., 2005). The plant homeodomain finger protein (PHF14) gene that is mutated in HCT116 cells contains a plant homeodomain (PHD) finger domain known to be involved in chromatin-mediated transcriptional regulation (Aasland *et al*., 1995). Although some

of these genes may not contribute to colon tumorigenesis directly, they may regulate the expression of known or yet unknown oncogenes or tumor suppressor genes that function downstream of these mutated genes. Functional *in vivo* studies in mice are required to provide a definite proof for the role of mutant genes in colon cancer development.

Only one mutant gene (*SMAD4* in HT29 cells) has been identified in two MSI-negative colon cancer cell lines HT29 and SW480 using the same GINI2 strategy. This is in agreement with the lower mutation rate reported in cells with functional MMR. It is noteworthy that both of these cell lines have *p53* mutations, whereas MSI-positive cell lines usually have wild-type *p53*. It is possible, therefore, that the higher frequency of detectable mutant genes in cells with inactivated MMR can compensate the absence of *p53* mutations. It is also possible that epigenetic mechanisms of gene inactivation may substitute for the mutational inactivation of genes in MSI-negative cells.

Methods

Cell culture and drugs treatments
Colon cancer cell lines were obtained from American Type Culture Collection repository and maintained in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum and antibiotics. All NMD inhibiting drugs were added when cells achieved 50–90% of confluence.

The emetine treatment (*GINI1*) protocol was as described previously (Ionov *et al*., 2004b). Briefly, cells were incubated for 8 h in tissue culture medium containing 100 μg/ml of emetine before total RNA was extracted from emetine treated or untreated cells. The caffeine, cantharidin and okadaic acid treatment protocols were as described for emetine using concentrations of 10 mM for caffeine and 100 nM for okadaic acid or cantharidin. In the latter two cases, the incubation times were reduced to 3 h to reduce cell death.

The GINI2 treatment was as follows. Cells were seeded in two tissue culture plates and caffeine (10 mM) was added to both plates. Following 4-h incubation, the medium from both plates was removed, cells were washed twice with phosphate-buffered saline and actinomycin D (2 μg/ml) together with caffeine (10 mM) was added to one plate and actinomycin D alone was added to the other plate. Following further 4-h incubation, total RNA from both plates was prepared and used for Northern blotting as well as for Affymetrix U133Plus2.0 oligonucleotide array analysis.

RNA isolation and Northern blot
Total RNAs for microarray and Northern blot analysis were prepared using TRIZOL reagents (Invitrogen, Carlsbad, CA, USA). For Northern blots, 20 μg of total RNA was size fractionated on 1% agarose formaldehyde gels and then transferred to nylon membranes (Hybond-N, Amersham, Boston, MA, USA). Hybridizations with radiolabeled probes were carried out at 68°C using MyracleHyb solution (Stratagene, Cedar Creek, TX, USA) according to the manufacturer's protocol. Radioactivity was detected and quantified using a PhosphorImager (Molecular Dynamics, Sunnyvale, CA, USA). Radiolabelled probes generated by PCR from RNA-derived cDNA were generated using the random-primed DNA labeling kit (Roche, Indianapolis, IN, USA).

Affymetrix oligonucleotide array data analysis
Messenger RNA levels in drug treated and untreated cells were measured using the Affymetrix U133Plus2.0 array. GeneChip expression array analysis was performed using Affymetrix Microarray Suite software version 5 (GeneChip Analysis Suite). This software uses two independent sets of algorithms, a quantitative algorithm (robust estimator of the mean difference in probe intensities) that computes the raw signal and the SLR directly from the hybridization intensities of the probes, as well as a confidence algorithm (nonparametric) that provides *P* values to estimate the confidence in detection (absent or present) or change (increase, decrease or no change) of a specific target.

An adaptation of RMA (Irizarry *et al*., 2003), with specific correction for GC biases known as GC-RMA (Wu *et al*., 2004) procedure was applied for the .CEL files of Affymetrix U133Plus2.0 data to further filter candidates for sequencing analysis. The resultant SLRs were generated based on pairwise comparison between the various treatments and control groups. GC-RMA uses only the perfect match (PM) intensity values and ignores the mismatch (MM) intensities which have been shown to introduce variation (Naef *et al*., 2002). The normalized PM values were then log transformed and all the probes in a set representing specific genes were analysed using Tukey's median polishing procedure.

Sequencing analysis
One microgram of total RNA from emetine treated colon cancer cells was reverse transcribed using the SuperScript II protocol (Invitrogen, Carlsbad, CA, USA). Overlapping PCR primer sets were used to generate products spanning the entire open reading frames for candidate genes. Primers for sequencing analysis were designed using Primer3 software available online (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi). Genomic DNA samples from colon tumors with MSI were purchased from Oncomatrix (Ocean Side, CA, USA). Primers for PCR amplification of genomic DNA fragments containing coding microsatellite sequences were designed using Primer3 software and are available upon request. The PCR products were gel purified and sequenced using the Applied Biosystems' PRISM 3100 Genetic Analyzer.

Analysis of frameshift mutations in DNA from primary colon tumors
Regions of approximately 100 bp encompassing the mononucleotide repeats of specific genes were amplified by PCR using specific primers (sequences are available upon request). PCR was carried out with Vent DNA polymerase (New England Biolabs, Boston, MA, USA) with one cycle at 94°C for 4 min followed by 30 cycles of 94, 54 and 68°C, for 30 s in each cycle in the presence of 0.2 μCi of [^{33}P]dCTP (1 Ci = 37 GBq) (PerkinElmer, Boston, MA, USA). PCR products were electrophoresed in a SequaGel XR (National Diagnostics, Atlanta, Georgia, USA) polyacrylamide gel. The gel was dried on filter paper and subjected to autoradiography.

Acknowledgements

This work was supported by grants Roswell Park Alliance Foundation grant, 1R01 CA109256-01 from the National Cancer Institute and PC030076 from the Prostate Program of the United States Army Department of Defense (Yurij Ionov) and in part by the RPCI Cancer Center Support Grant CA 16056 for all authors and NIH NSO54543 (John Cowell).

References

Aaltonen LA, Peltomaki P, Leach FS, Sistonen P, Pylkkanen L, Mecklin JP *et al*. (1993). Clues to the pathogenesis of familial colorectal cancer. *Science* **260**: 812–816.

Aasland R, Gibson TJ, Stewart AF. (1995). The PHD finger: implications for chromatin-mediated transcriptional regulation. *Trends Biochem Sci* **20**: 56–59.

Boland CR, Thibodeau SN, Hamilton SR, Sidransky D, Eshleman JR, Burt RW *et al*. (1998). A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. *Cancer Res* **58**: 5248–5257.

Brumbaugh KM, Otterness DM, Geisen C, Oliveira V, Brognard J, Li X *et al*. (2004). The mRNA surveillance protein hSMG-1 functions in genotoxic stress response pathways in mammalian cells. *Mol Cell* **14**: 585–598.

Chen J, Heerdt BG, Augenlicht LH. (1995). Presence and instability of repetitive elements in sequences the altered expression of which characterizes risk for colonic cancer. *Cancer Res* **55**: 174–180.

Duval A, Reperant M, Hamelin R. (2002). Comparative analysis of mutation frequency of coding and non coding short mononucleotide repeats in mismatch repair deficient colorectal cancers. *Oncogene* **21**: 8062–8066.

Duval A, Rolland S, Compoint A, Tubacher E, Iacopetta B, Thomas G *et al*. (2001). Evolution of instability at coding and non-coding repeat sequences in human MSI-H colorectal cancers. *Hum Mol Genet* **10**: 513–518.

El-Bchiri J, Buhard O, Penard-Lacronique V, Thomas G, Hamelin R, Duval A. (2005). Differential nonsense mediated decay of mutated mRNAs in mismatch repair deficient colorectal cancers. *Hum Mol Genet* **14**: 2435–2442.

Fang S, Lorick KL, Jensen JP, Weissman AM. (2003). RING finger ubiquitin protein ligases: implications for tumorigenesis, metastasis and for molecular targets in cancer. *Semin Cancer Biol* **13**: 5–14.

Fishel R, Lescoe MK, Rao MR, Copeland NG, Jenkins NA, Garber J *et al*. (1993). The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. *Cell* **75**: 1027–1038.

Garnon J, Lachance C, Di Marco S, Hel Z, Marion D, Ruiz MC *et al*. (2005). Fragile X-related protein FXR1P regulates proinflammatory cytokine tumor necrosis factor expression at the post-transcriptional level. *J Biol Chem* **280**: 5750–5763.

Girault I, Lerebours F, Amarir S, Tozlu S, Tubiana-Hulin M, Lidereau R *et al*. (2003). Expression analysis of estrogen receptor alpha coregulators in breast carcinoma: evidence that NCOR1 expression is predictive of the response to tamoxifen. *Clin Cancer Res* **9**: 1259–1266.

Holbrook JA, Neu-Yilik G, Hentze MW, Kulozik AE. (2004). Nonsense-mediated decay approaches the clinic. *Nat Genet* **36**: 801–808.

Huusko P, Ponciano-Jackson D, Wolf M, Kiefer JA, Azorsa DO, Tuzmen S *et al*. (2004). Nonsense-mediated decay microarray analysis identifies mutations of EPHB2 in human prostate cancer. *Nat Genet* **36**: 979–983.

Ionov Y, Matsui S, Cowell JK. (2004a). A role for p300/CREB binding protein genes in promoting cancer progression in colon cancer cell lines with microsatellite instability. *Proc Natl Acad Sci USA* **101**: 1273–1278.

Ionov Y, Nowak N, Perucho M, Markowitz S, Cowell JK. (2004b). Manipulation of nonsense mediated decay identifies gene mutations in colon cancer Cells with microsatellite instability. *Oncogene* **23**: 639–645.

Ionov Y, Peinado MA, Malkhosyan S, Shibata D, Perucho M. (1993). Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. *Nature* **363**: 558–561.

Ireton RC, Davis MA, van Hengel J, Mariner DJ, Barnes K, Thoreson MA *et al*. (2002). A novel role for p120 catenin in E-cadherin function. *J Cell Biol* **159**: 465–476.

Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U *et al*. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**: 249–264.

Ishigaki Y, Li X, Serin G, Maquat LE. (2001). Evidence for a pioneer round of mRNA translation: mRNAs subject to nonsense-mediated decay in mammalian cells are bound by CBP80 and CBP20. *Cell* **106**: 607–617.

Keeton EK, Brown M. (2005). Cell cycle progression stimulated by tamoxifen-bound estrogen receptor-alpha and promoter-specific effects in breast cancer cells deficient in N-CoR and SMRT. *Mol Endocrinol* **19**: 1543–1554.

Knudson Jr AG. (1971). Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci USA* **68**: 820–823.

Konishi M, Kikuchi-Yanoshita R, Tanaka K, Muraoka M, Onda A, Okumura Y *et al*. (1996). Molecular nature of colon tumors in hereditary nonpolyposis colon cancer, familial polyposis, and sporadic colon cancer. *Gastroenterology* **111**: 307–317.

Kunkel TA, Patel SS, Johnson KA. (1994). Error-prone replication of repeated DNA sequences by T7 DNA polymerase in the absence of its processivity subunit. *Proc Natl Acad Sci USA* **91**: 6830–6834.

Leach FS, Nicolaides NC, Papadopoulos N, Liu B, Jen J, Parsons R *et al*. (1993). Mutations of a mutS homolog in hereditary nonpolyposis colorectal cancer. *Cell* **75**: 1215–1225.

Lengauer C, Kinzler KW, Vogelstein B. (1998). Genetic instabilities in human cancers. *Nature* **396**: 643–649.

Liu L, Amy V, Liu G, McKeehan WL. (2002). Novel complex integrating mitochondria and the microtubular cytoskeleton with chromosome remodeling and tumor suppressor RASSF1 deduced by in silico homology analysis, interaction cloning in yeast, and colocalization in cultured cells. *In vitro Cell Dev Biol Anim* **38**: 582–594.

Liu L, Vo A, Liu G, McKeehan WL. (2005a). Distinct structural domains within C19ORF5 support association with stabilized microtubules and mitochondrial aggregation and genome destruction. *Cancer Res* **65**: 4191–4201.

Liu L, Vo A, Liu G, McKeehan WL. (2005b). Putative tumor suppressor RASSF1 interactive protein and cell death inducer C19ORF5 is a DNA binding protein. *Biochem Biophys Res Commun* **332**: 670–676.

Liu L, Vo A, McKeehan WL. (2005c). Specificity of the methylation-suppressed A isoform of candidate tumor suppressor RASSF1 for microtubule hyperstabilization is determined by cell death inducer C19ORF5. *Cancer Res* **65**: 1830–1838.

Maquat LE. (2005). Nonsense-mediated mRNA decay in mammals. *J Cell Sci* **118**: 1773–1776.

Markowitz S, Wang J, Myeroff L, Parsons R, Sun L, Lutterbaugh J *et al*. (1995). Inactivation of the type II TGF-beta receptor in colon cancer cells with microsatellite instability. *Science* **268**: 1336–1338.

Naef F, Hacker CR, Patil N, Magnasco M. (2002). Characterization of the expression ratio noise structure in high-density oligonucleotide arrays. *Genome Biol* **3**: research0018.1–0018.11.

Noensie EN, Dietz HC. (2001). A strategy for disease gene identification through nonsense-mediated mRNA decay inhibition. *Nat Biotechnol* **19**: 434–439.

Ohnishi T, Yamashita A, Kashima I, Schell T, Anders KR, Grimson A *et al*. (2003). Phosphorylation of hUPF1 induces formation of mRNA surveillance complexes containing hSMG-5 and hSMG-7. *Mol Cell* **12**: 1187–1200.

Oliveira C, Velho S, Domingo E, Preto A, Hofstra RM, Hamelin R *et al*. (2005). Concomitant RASSF1A hypermethylation and KRAS/BRAF mutations occur preferentially in MSI sporadic colorectal cancer. *Oncogene* **24**: 7630–7634.

Perucho M. (1999a). Correspondence re: C.R. Boland et al., A National Cancer Institute workshop on microsatellite instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. *Cancer Res* **58**: 5248–5257.

Perucho M. (1999b). Correspondence re: C.R. Boland et al., A National Cancer Institute workshop on microsatellite instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. *Cancer Res* **59**: 249–56.

Rampino N, Yamamoto H, Ionov Y, Li Y, Sawai H, Reed JC, *et al*. (1997). Somatic frameshift mutations in the BAX gene in colon cancers of the microsatellite mutator phenotype. *Science* **275**: 967–969.

Rossi MR, Hawthorn L, Platt J, Burkhardt T, Cowell JK, Ionov Y. (2005). Identification of inactivating mutations in the JAK1, SYNJ2, and CLPTM1 genes in prostate cancer cells using inhibition of nonsense-mediated decay and microarray analysis. *Cancer Genet Cytogenet* **161**: 97–103.

Salovaara R, Roth S, Loukola A, Launonen V, Sistonen P, Avizienyte E *et al*. (2002). Frequent loss of SMAD4/DPC4 protein in colorectal cancers. *Gut* **51**: 56–59.

Suzuki K, Dai T, Suzuki I, Dai Y, Yamashita K, Perucho M. (2002). Low mutation incidence in polymorphic noncoding short mononucleotide repeats in gastrointestinal cancer of the microsatellite mutator phenotype pathway. *Cancer Res* **62**: 1961–1965.

Thibodeau SN, Bren G, Schaid D. (1993). Microsatellite instability in cancer of the proximal colon. *Science* **260**: 816–819.

Thoreson MA, Reynolds AB. (2002). Altered expression of the catenin p120 in human cancer: implications for tumor progression. *Differentiation* **70**: 583–589.

Usuki F, Yamashita A, Higuchi I, Ohnishi T, Shiraishi T, Osame M *et al*. (2004). Inhibition of nonsense-mediated mRNA decay rescues the phenotype in Ullrich's disease. *Ann Neurol* **55**: 740–744.

Veigl ML, Kasturi L, Olechnowicz J, Ma AH, Lutterbaugh JD, Periyasamy S *et al*. (1998). Biallelic inactivation of hMLH1 by epigenetic gene silencing, a novel mechanism causing human MSI cancers. *Proc Natl Acad Sci USA* **95**: 8698–8702.

Vogelstein B, Kinzler KW. (2004). Cancer genes and the pathways they control. *Nat Med* **10**: 789–799.

Weischenfeldt J, Lykke-Andersen J, Porse B. (2005). Messenger RNA surveillance: neutralizing natural nonsense. *Curr Biol* **15**: R559–62.

Woerner SM, Benner A, Sutter C, Schiller M, Yuan YP, Keller G *et al*. (2003). Pathogenesis of DNA repair-deficient cancers: a statistical meta-analysis of putative Real Common Target genes. *Oncogene* **22**: 2226–2235.

Woodford-Richens KL, Rowan AJ, Gorman P, Halford S, Bicknell DC, Wasan HS *et al*. (2001). SMAD4 mutations in colorectal cancer probably occur before chromosomal instability, but after divergence of the microsatellite instability pathway. *Proc Natl Acad Sci USA* **98**: 9719–9723.

Wu Z, Irizarry RA, Gentleman R, Murillo FM, Spencer F. (2004). A model based background adjustment for oligonucleotide expression arrays (May 28, 2004). Johns Hopkins University, Department of Biostatistics Working Papers. Working Paper 1.

Zhang L, Yu J, Willson JK, Markowitz SD, Kinzler KW, Vogelstein B. (2001). Short mononucleotide repeat sequence variability in mismatch repair-deficient cancers. *Cancer Res* **61**: 3801–3805.

Zhang Z, Yamashita H, Toyama T, Sugiura H, Ando Y, Mita K *et al*. (2005). NCOR1 mRNA is an independent prognostic factor for breast cancer. *Cancer Lett* **237**: 123–129.

Supporting Data:

Figure 1 - Figure 7

Figure 1: Chromosome 22q profiles from Pollack J.R. *et al.*, 2002 (7)

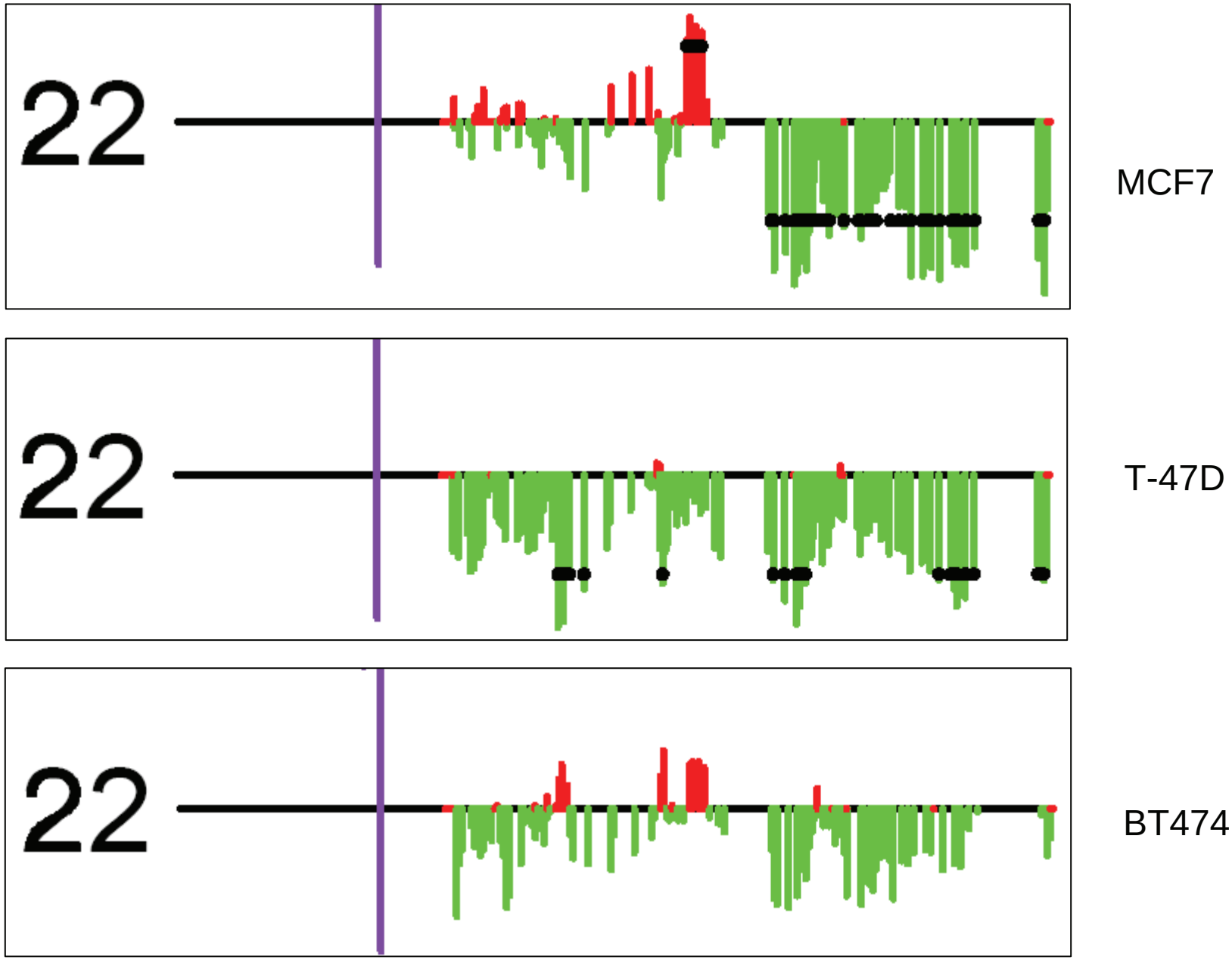


Fig 2A

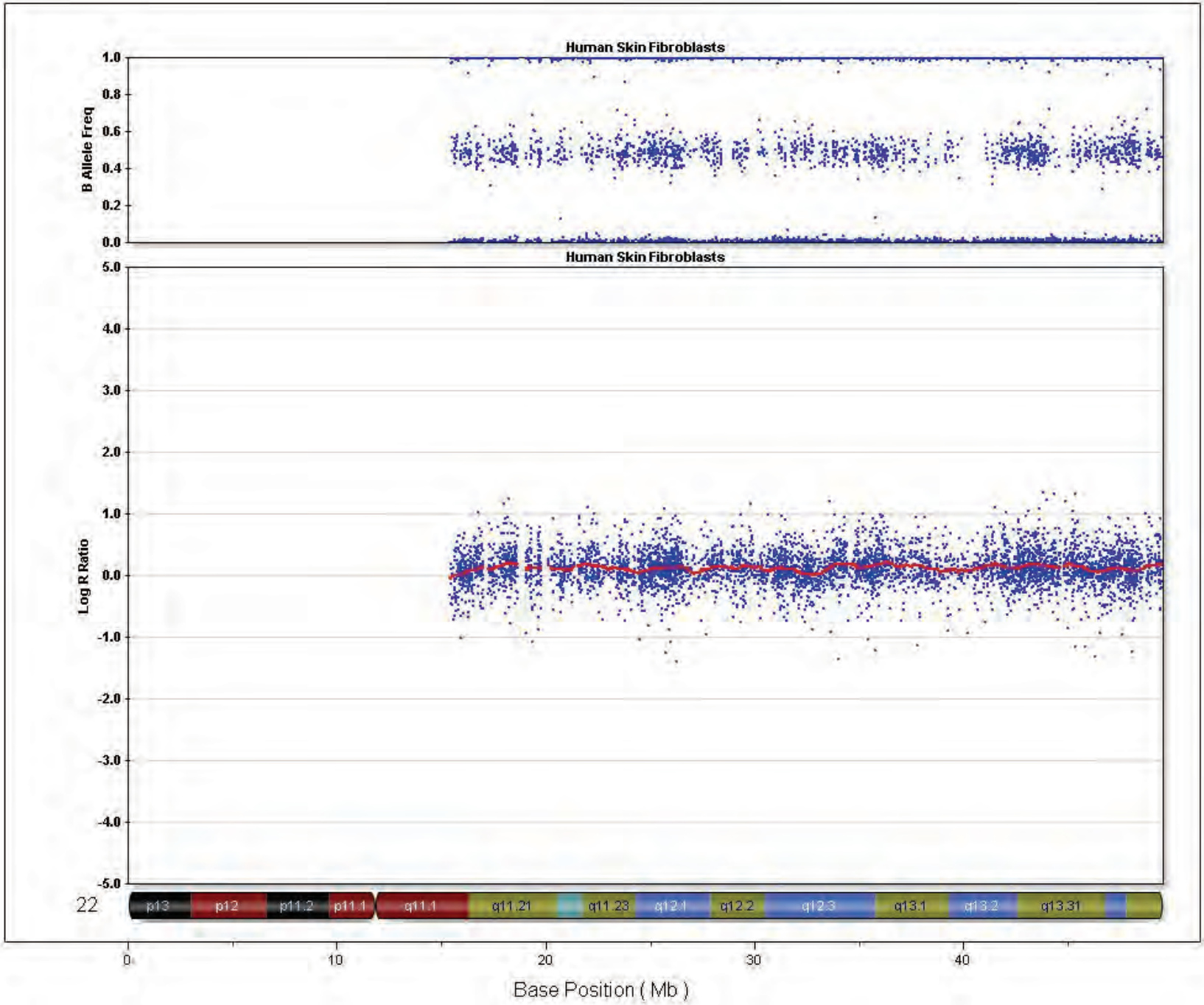

Human Skin Fibroblasts
B Allele Freq
1.0
0.8
0.6
0.4
0.2
0.0
Human Skin Fibroblasts
Log R Ratio
5.0
4.0
3.0
2.0
1.0
0.0
-1.0
-2.0
-3.0
-4.0
-5.0
22
p13
p12
p11.2
p11.1
q11.1
q11.21
q11.23
q12.1
q12.2
q12.3
q13.1
q13.2
q13.31
0
10
20
30
40
Base Position (Mb)

Fig 2B

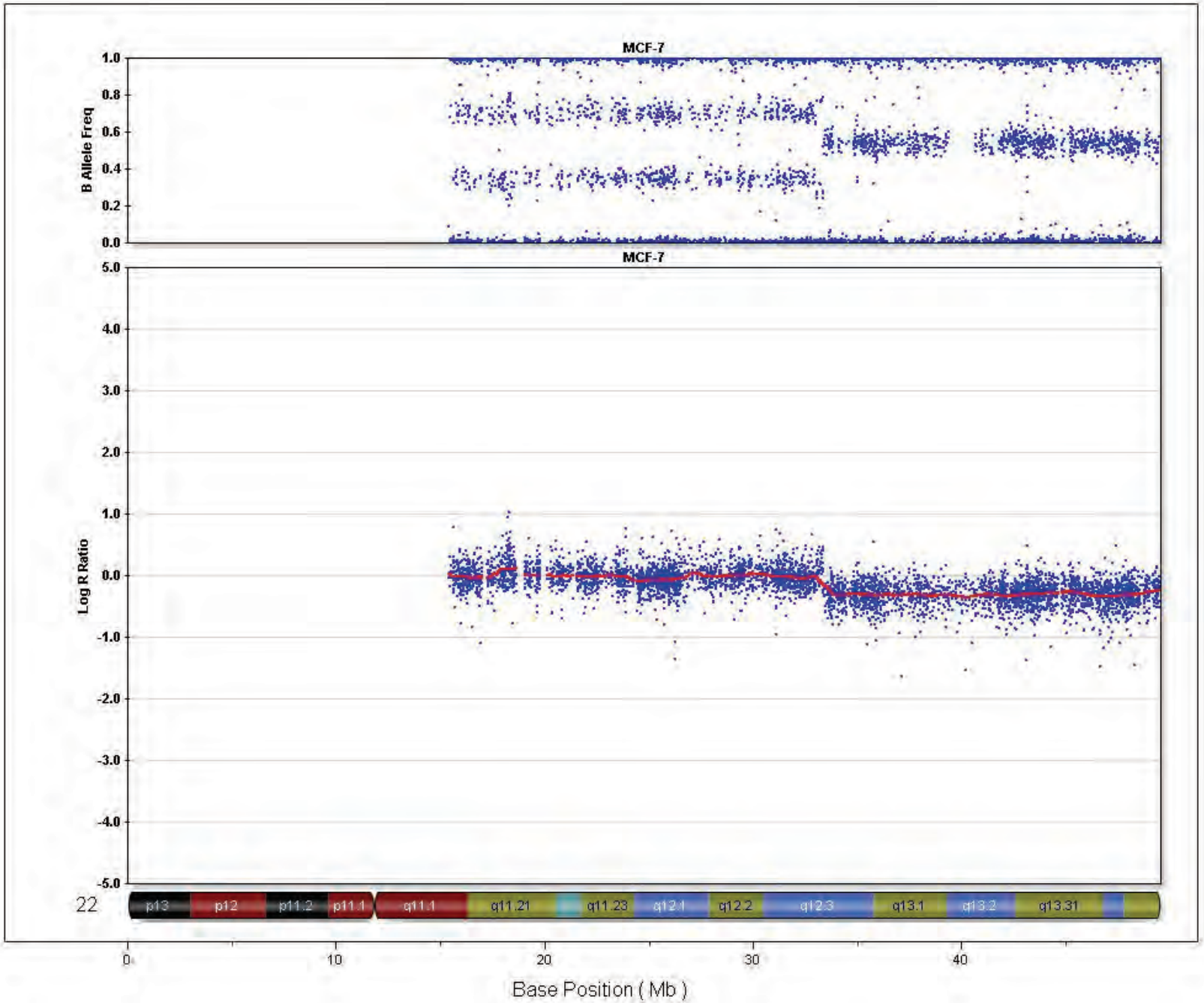

MCF-7
B Allele Freq
1.0
0.8
0.6
0.4
0.2
0.0
MCF-7
Log R Ratio
5.0
4.0
3.0
2.0
1.0
0.0
-1.0
-2.0
-3.0
-4.0
-5.0
22
p13
p12
p11.2
p11.1
q11.1
q11.21
q11.23
q12.1
q12.2
q12.3
q13.1
q13.2
q13.31
0
10
20
30
40
Base Position (Mb)

Fig 2C

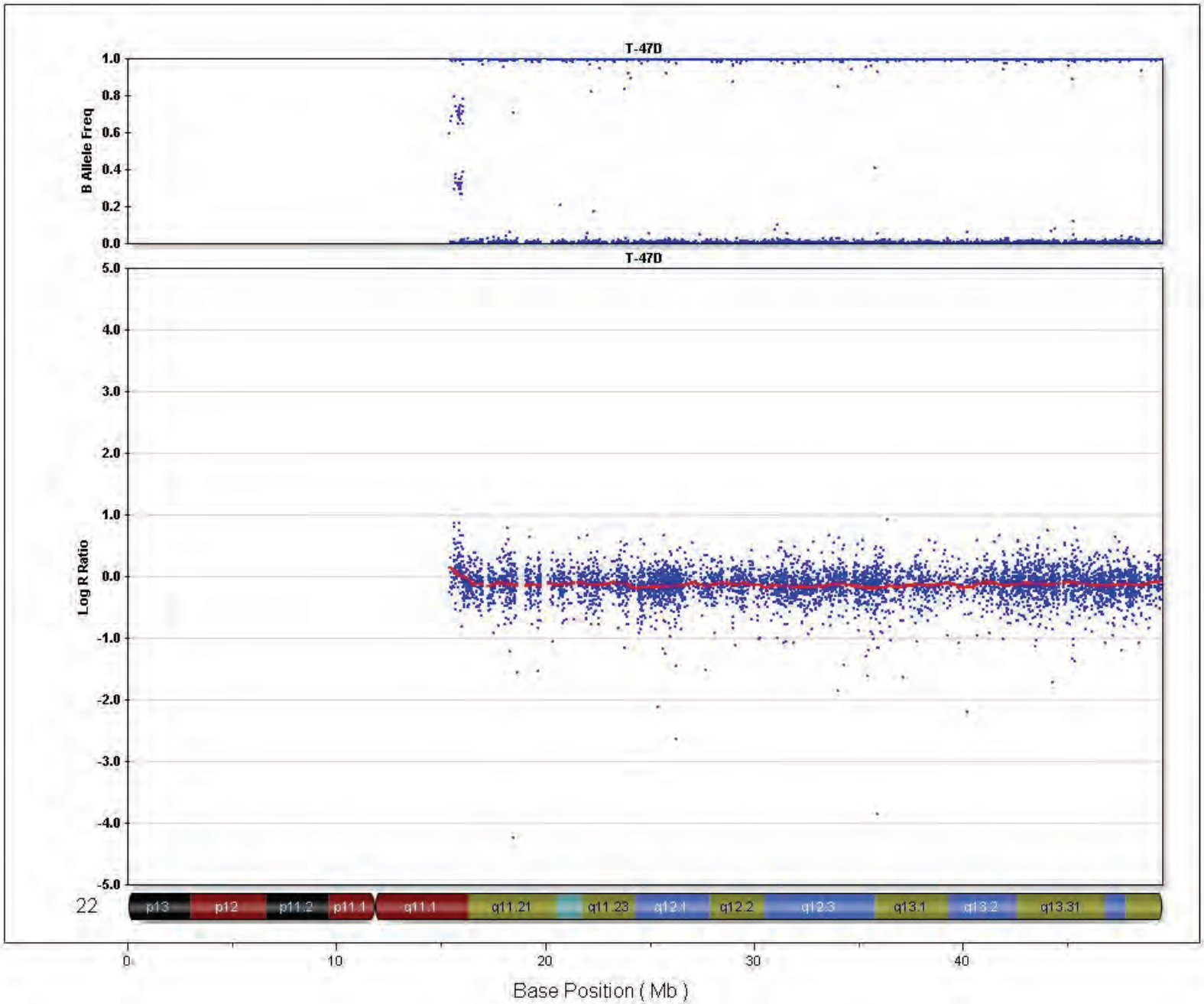

T-47D
B Allele Freq
1.0
0.8
0.6
0.4
0.2
0.0
T-47D
Log R Ratio
5.0
4.0
3.0
2.0
1.0
0.0
-1.0
-2.0
-3.0
-4.0
-5.0
22
p13
p12
p11.2
p11.1
q11.1
q11.21
q11.23
q12.1
q12.2
q12.3
q13.1
q13.2
q13.31
0
10
20
30
40
Base Position (Mb)

Fig 2D

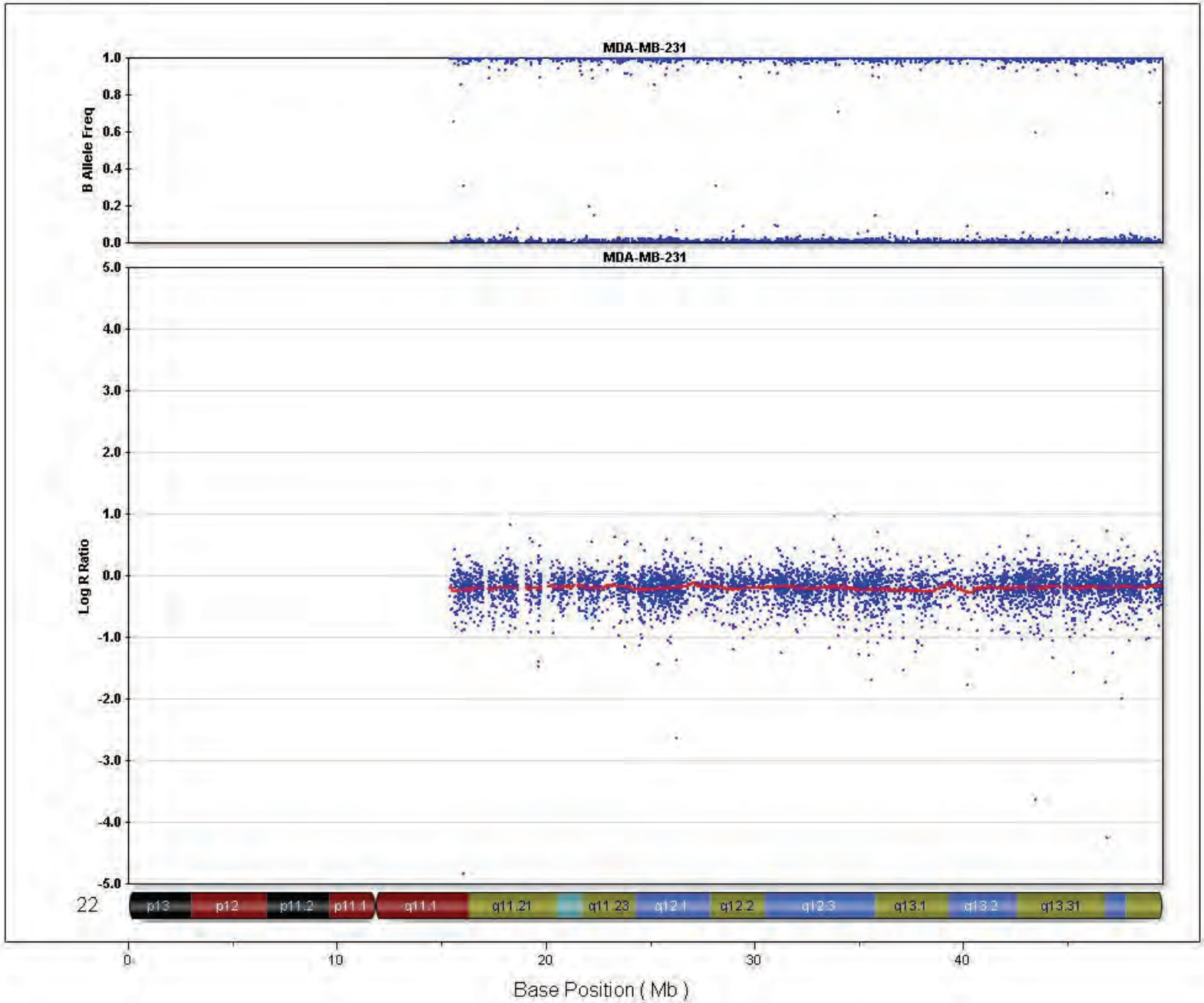
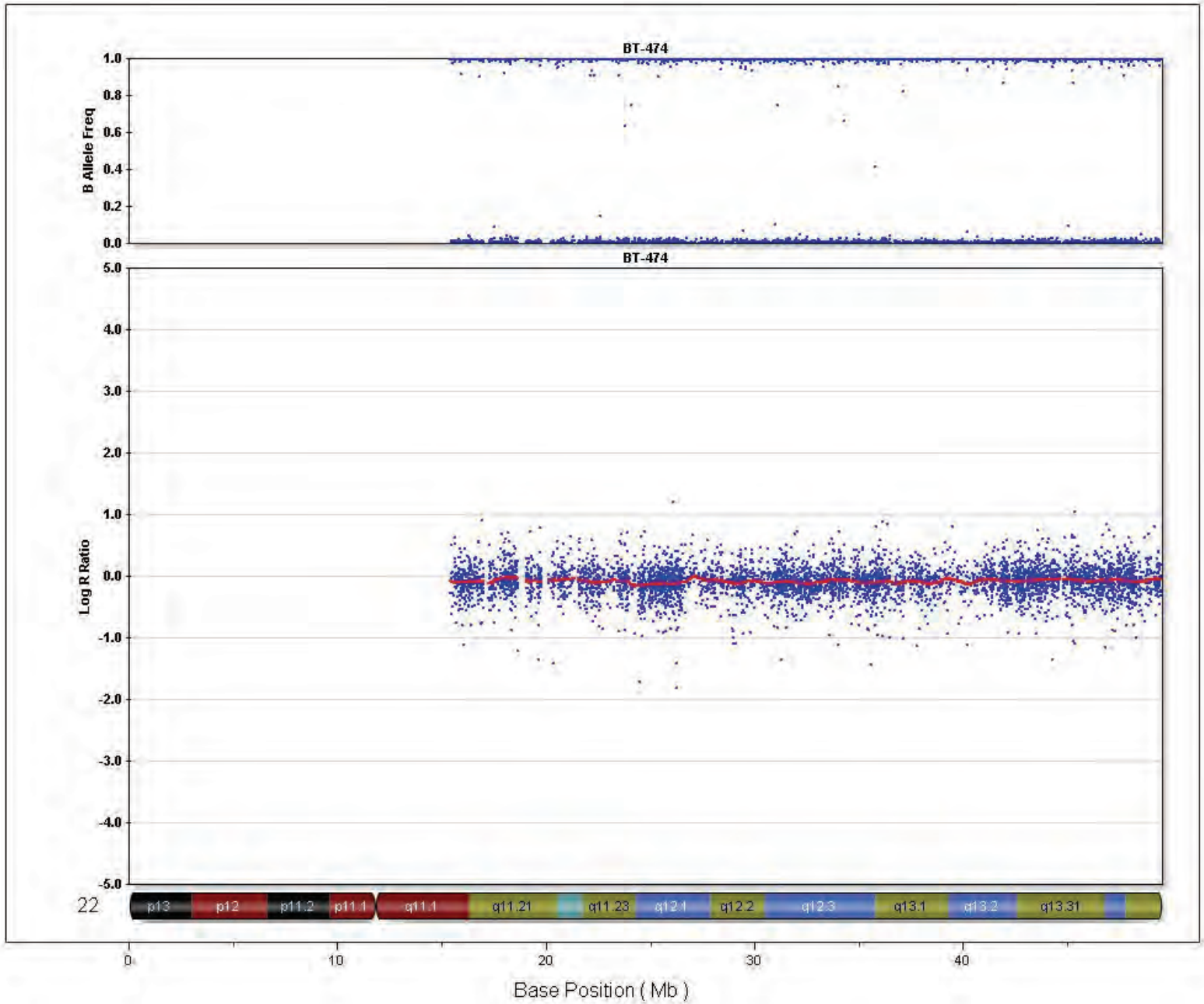

MDA-MB-231
B Allele Freq
1.0
0.8
0.6
0.4
0.2
0.0
MDA-MB-231
Log R Ratio
5.0
4.0
3.0
2.0
1.0
0.0
-1.0
-2.0
-3.0
-4.0
-5.0
22
p13
p12
p11.2
p11.1
q11.1
q11.21
q11.23
q12.1
q12.2
q12.3
q13.1
q13.2
q13.31
0
10
20
30
40
Base Position (Mb)

Fig 2E


BT-474
B Allele Freq
1.0
0.8
0.6
0.4
0.2
0.0
BT-474
Log R Ratio
5.0
4.0
3.0
2.0
1.0
0.0
-1.0
-2.0
-3.0
-4.0
-5.0
22
p13
p12
p11.2
p11.1
q11.1
q11.21
q11.23
q12.1
q12.2
q12.3
q13.1
q13.2
q13.31
0
10
20
30
40
Base Position (Mb)

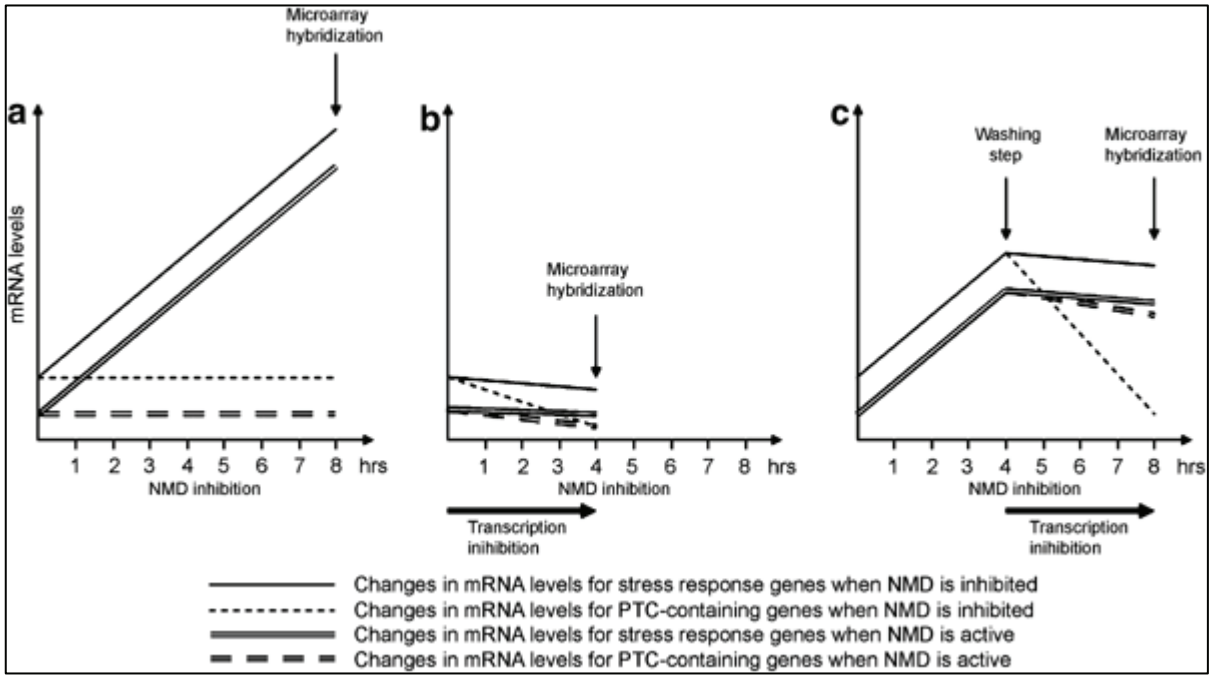


Figure 3: Adapted from Ivanov I. et al., 2007 (5). Theoretical depiction of changes in mRNA levels for stress response and PTC-containing genes following different protocols to inhibit NMD. **(a)** Continuous treatment of cells with an inhibitor of NMD results in increase of mRNA levels for PTC-containing genes, as well as for stress response genes. Microarray hybridization analysis cannot distinguish between the two causes of these increases: the enhanced mRNA synthesis due to stress response or the decreased mRNA degradation in case of PTC-containing genes. **(b)** When transcription is prevented using Actinomycin D, even though NMD is blocked, the levels of PTC-containing transcripts and stress response genes are too low to be efficiently detected by microarrays. Changes in mRNA levels induced by NMD inhibition detected by hybridization analysis for many genes are within the 'hybridization noise', and are not reliable. **(c)** Blocking NMD for 4 h allows the accumulation of both PTC-containing transcripts and stress response gene transcripts to levels where they should be detected on the expression arrays. When transcription is blocked after this initial accumulation, the degradation of PTC-containing transcripts should occur quickly if the NMD blocking agent is removed, or will be delayed when NMD is sustained. The degradation of stress response transcripts after blocking transcription, on the other hand, does not depend on NMD. As a result, significant mRNA differences should be detected on the expression array for PTC-containing genes but not for the stress response genes.

Figure 4: Comparison of GINI-1 and GINI-2 for ability to detect a *TP53* nonsense mutation in PC-3 prostate cancer cells

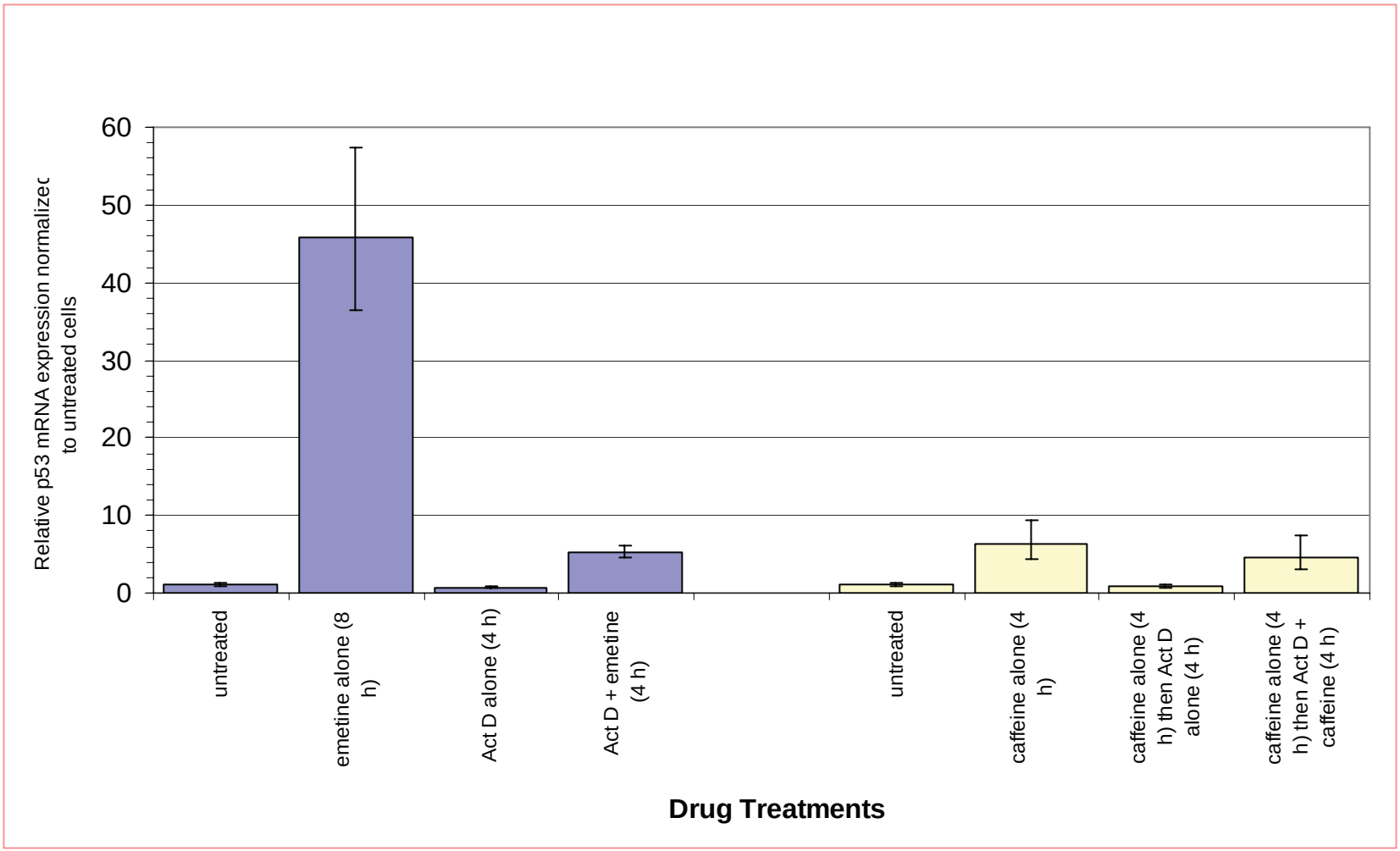


Figure 5: Examination of off-target effects of GINI-1 and GINI-2: modulation of wild-type p53 mRNA levels in MCF-7 cells

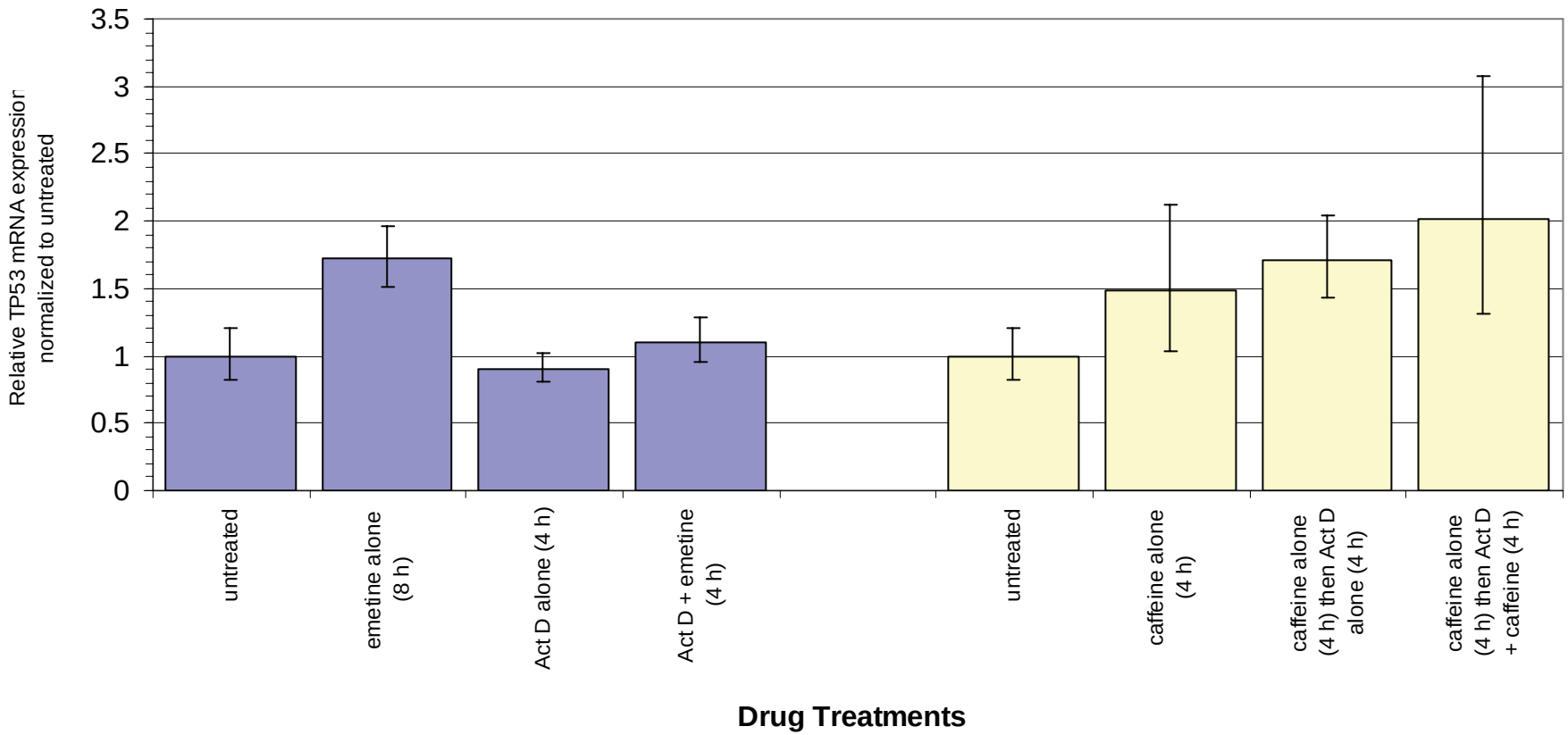


Figure 6: Examination of off-target effects of GINI-1 and GINI-2: modulation of wild-type total PPARgamma mRNA levels in PC-3 cells

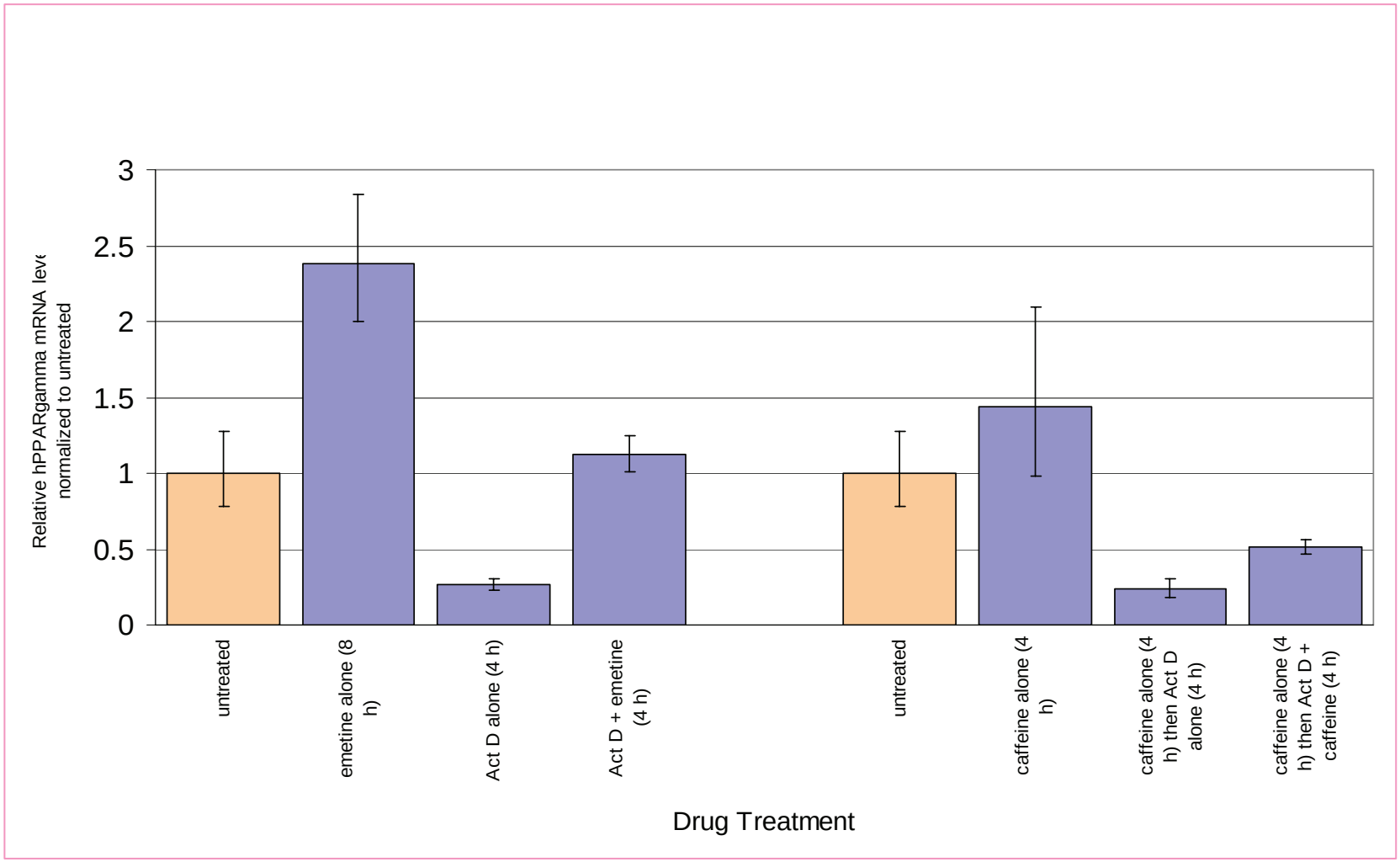


Figure 7: Examination of off-target effects of GINI-1 and GINI-2: modulation of wild-type total PPARgamma mRNA levels in MCF-7 cells